Measuring Individual Treatment Benefits Using Longitudinal Outcomes from Clinical Trials or Hospital Data

By

Xuan Zhang
M.S., University of Kansas, 2016
Ph.D., Shanxi Medical University, 2001
M.B.B.S., Shanxi Medical University, 1996

Submitted to the graduate degree program in Biostatistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson: Dr. Francisco J. Diaz, Ph.D.

Dr. Jo Wick, Ph.D.

Dr. Jonathan Mahnken, Ph.D.

Dr. Milind Phadnis, Ph.D.

Dr. Mark Chertoff, Ph.D.

Date Defended: October 28, 2019

The dissertation committee for Xuan Zhang
certifies that this is the approved version of the following dissertation:

Measuring Individual Treatment Benefits Using Longitudinal Outcomes from Clinical Trials or Hospital Data

_____

Chairperson: Dr. Francisco Diaz, Ph.D.

_____

Biostatistics Graduate Director: Dr. Jo Wick, Ph.D.

Date Approved: Nov. 11, 2019

Abstract

It is increasingly recognized that a patient's response to a medical treatment is a statistically heterogeneous phenomenon. The average treatment effects may not represent a heterogeneous population of patients. The benefits each patient receive from the treatment could differ, requiring measurement of treatment benefits at the patient level. Despite of the development of methods in this field, new methods are needed for predicting individual treatment benefits using longitudinal binary outcomes or hospital data with nonignorable missingness.

This dissertation has three main chapters. Chapter 1 introduces a method for predicting individual treatment benefits based on a personalized medicine model that implements random effects logistic regression of binary outcomes that may change over time. The method uses empirical Bayes (EB) estimators based on patients' characteristics and responses to treatment. The prediction performance is evaluated in simulated new patients using correlations between the predicted and the true benefits as well as relative biases of the predicted benefits versus the true benefits. As an application, the method is used to examine changes in the disorganized dimension of antipsychotic-naïve patients from an antipsychotic randomized clinical trial.

Chapter two of the dissertation presents a method for predicting individual treatment benefits with a novel 2-dimensional personalized medicine model that handles non-ignorable missingness due to hospital discharge and evaluate its reliability and accuracy by simulations. The longitudinal outcome of interest is modeled simultaneously with the hospital length of stay through a joint mixed model. The method is illustrated with an application assessing individual pain management benefits post spine fusion surgery. EB-Predicted individual benefits are compared with Monte-Carlo computed benefits. Pearson's correlations and relative biases are used to assess the prediction accuracy.

Finally, Chapter three of the dissertation applies the methodology developed in Chapter two to analyze with more clinical detail the impact of depression and age on individual benefits of postoperative pain management in lumbar spinal fusion patients using Cerner HealthFacts® electronic health records. The developed joint multivariate mixed model of pain scores and length of hospital stay is used to analyze individual benefits. The effects of depression and age on the amount and rate of change of the pain management benefits are evaluated, as well as the association between individual benefits and post-surgical hospital length of stay.

We conclude that the utilization of the EB prediction of individual treatment benefits is useful in the analyses of treatment effects using not only clinical trial data but also electronic health records. Predicted individual treatment benefits are accurate when model parameters are reliably estimated.

Acknowledgements

Many thanks go to my advisor Dr. Francisco J. Diaz, Ph.D. Without his guidance I would not be able to complete this dissertation.

I also thank our collaborators Dr. Nikos from the Department of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Medical School, Athens, Greece, for his advice regarding the theory and application of joint mixed effects models; Dr. Jose de Leon, M.D. from the Mental Health Research Center at Eastern State Hospital, Lexington, KY, for his constructive suggestions and support in the clinical analysis of the study results; and Dr. Benedicto Crespo-Facorro, M.D., from the Department of Psychiatry, Marqués de Valdecilla University Hospital, School of Medicine, University of Cantabria, Santander, Spain, for allowing us to use the dataset from his antipsychotic clinical trial in Chapter one.

Special thanks go to Boston Strategic Partners, Inc., Boston, MA, for allowing us to use the Cerner HealthFacts® electronic health records for Chapters two and three.

Finally, I thank my other committee members Drs. Jo Wick, Ph.D., Jonathan Mahnken, Ph.D., Milind Phadnis, Ph.D., and Mark Chertoff, Ph.D. for their valuable comments.

Here are the acknowledgments and funding sources for the study of Chapter 1:

The antipsychotic data was collected in a randomized clinical trial supported by several grants described in Pelayo-Teran et al. (2014). Drs. Diaz, Zhang and de Leon were not involved in the funding, planning or operation of the RCT. Dr. Diaz was supported by an Institutional Clinical and Translational Science Award from the National Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366 (KL2/TL1), awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH. The authors thank Drs. Jonathan

Mahnken, Jo Wick and Milind Phadnis for their comments and suggestions, and Lorraine Maw, M.A., who helped in editing this article.

Here are the acknowledgments and funding sources for the study of Chapter 2:

Dr. Diaz was supported in part by an Institutional Clinical and Translational Science Award from the National Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366, awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH or NCATS. The authors thank Drs. Jonathan Mahnken, Jo Wick, and Milind Phadnis for their comments and suggestions.

Here are the acknowledgments and funding sources for the study of Chapter 3:

Dr. Diaz was supported in part by an Institutional Clinical and Translational Science Award from the National Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366, awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH or NCATS. The authors thank Lorraine Maw, M.A., from the Mental Health Research Center at Eastern State Hospital, Lexington, KY, USA, for editorial assistance. The authors thank Drs. Jonathan Mahnken, Jo Wick, and Milind Phadnis for their comments and suggestions.

Table of Contents

Chapter 1: Measuring individual benefits of psychiatric treatment using longitudinal binary outcomes: Application to antipsychotic benefits in non-cannabis and cannabis users

In collaboration with Drs. Benedicto Crespo-Facorro, M.D., Jose de Leon, M.D., and Francisco J. Diaz, Ph.D.

(See collaborator affiliations in Acknowledgements)

1.0. Abstract

We present and evaluate a method for predicting individual treatment benefits based on random effects logistic regression models of binary outcomes that change over time. The method uses empirical Bayes estimators based on patients' characteristics and responses to treatment. It is applicable to both 1-dimentional and 2-dimentional personalized medicine models. Comparisons between predicted and true benefits of simulated new patients using correlations and relative biases were used to evaluate prediction performance. The predicted benefits had relatively small relative biases and relatively high correlations with the true benefits in the simulated new patients. The predictors also captured overall population trends in the evolution of individual benefits. The proposed approach can be used to retrospectively evaluate patients' responses in a clinical trial, or to retrospectively or prospectively predict individual benefits of different treatments for new patients using patients' characteristics and previous responses. The method is used to examine changes in the disorganized dimension of antipsychotic-naïve patients from an antipsychotic randomized clinical trial. Retrospective prediction of individual benefits revealed that more cannabis users had slower and lower responses to antipsychotic treatment as compared to non-cannabis users, revealing cannabis use as a negative prognostic factor for psychotic disorders in the disorganized dimension.

Keywords: individual benefits, longitudinal binary outcomes, cannabis, psychosis, empirical Bayesian prediction

1.1 Introduction

Randomized clinical trials (RCT) are often used to establish the best treatment for the average patient. Heterogeneity in patients' responses are largely overlooked. In the era of personalized medicine, which perceives patients' responses to medications as a heterogeneous phenomenon (de Leon, 2012; Ruberg et al., 2010; Sies et al., 2019; Xu and Hedeker, 2001), it is essential to develop statistical tools for the analysis of individualized treatment benefits in RCTs that guide therapy in medical practice.

The statistical approaches developed in recent years for establishing personalized treatment rules or predicting individualized treatment benefits include methods based on generalized linear mixed-effects models (Andrews and Cho, 2018; Botts et al., 2008; Cho 2017; Diaz, 2016, 2017; Diaz et al., 2007, 2008, 2012a, 2012b, 2013a, 2013b, 2014; Senn, 2016; Zhu and Qu, 2016), penalized regression for high-dimensional data (Boulesteix et al., 2017; Kim et al., 2017; Ma et al., 2016), and machine learning methods (Goldstein et al., 2017; Powers et al., 2018). Among them, generalized linear mixed-effects modeling is an excellent tool for predicting individuated treatment benefits. Diaz (2016, 2019) proposed the concepts of 1-dimensional personalized medicine (1-PM) and 2-dimensional personalized medicine (2-PM) models for treatments of chronic diseases using mixed effects. These models use random effects in addition to fixed effects to represent the heterogeneity of patients' characteristics including unknown traits. If the random effects only include a random intercept representing unexplained patient variability before treatment administration, then the model is considered a 1-PM model. In this case, the treatment effect is fixed in the sense that it is independent of the patient. If the random effects additionally include random coefficients whose variabilities are explained by differences in the treatment effect across patients, then it is considered a 2-PM model (Diaz 2016, 2019). While treatment effects are still measured with regression coefficients other than the intercept, the individual treatment benefit is a dimension that may also depend on some of the patient's known or unknown baseline characteristics.

Cannabis is a psychoactive drug widely used around the world; it has a significant impact on mental and physical health (Barrigón et al., 2010; Cobo et al., 2017; Legleye, 2018). Cannabis use has been shown to be associated with increased risk of developing psychotic disorders as well as adverse outcomes in

patients with psychosis (Linszen et al., 1994; Moore et al., 2007; Zammit et al., 2008). Multiple cohort studies suggest that cannabis abuse leads to more severe psychotic symptoms in patients with psychosis or schizophrenia (Caspari, 1999; Grech et al., 2005; Foti et al., 2010; Kuepper et al., 2011). Cannabis use is also known to be associated with increased relapse and non-adherence (Hides et al., 2006; Linszen et al., 1997; Schoeler, 2016). Clausen et al. (2014) found that patients who stopped using cannabis had a significantly lower level of psychotic symptoms after adjusting for baseline conditions and medications. In our pragmatic RCT of patients with a first episode of non-affective psychosis, after adjusting for potential confounders, cannabis use was associated with poorer responses to antipsychotic treatment when responses were measured with the disorganized or the positive dimensions of the Scale for the Assessment of Positive Symptoms and Negative Symptoms (SAPS-SANS) (Andreasen, 1983a, 1983b; Pelayo-Teran et al., 2014).

The objective of the current study is three-fold. The first is to extend the methodology for measuring individual treatment benefits proposed by Diaz (2016, 2019) to longitudinal binary outcomes, which utilizes empirical Bayesian (EB) predictors of individual benefits. This is necessary because patients' responses to treatment may change over time, whereas the previous approach to measuring individual benefits with binary outcomes considered only stable post-treatment responses (Diaz, 2016). The second is to evaluate the performance of the proposed EB predictors by showing that they correlate with the true benefits achieved by simulated hypothetical new patients and showing that the predictors can also reflect overall clinical population trends. The third is to illustrate the methodology by measuring the individual benefits of antipsychotic treatment and showing how cannabis use affects these, using the disorganized dimension scores of the SAPS-SANS scale from the patients of our pragmatic RCT (Pelayo-Teran et al., 2014).

In Section 2, we present the methods used for this study. Section 2.1 describes 1-PM and 2-PM logistic regression models for longitudinal responses that evolve over time. Section 2.2 describes disease severity measures and benefit functions under the logistic model. Section 2.3 describes how individual benefits can be calculated for each time point using EB prediction. Section 2.4 introduces the application of the proposed method of benefit prediction in the analysis of data from the antipsychotic RCT (Pelayo-Teran et al., 2014). Section 2.5 describes two methods for evaluating the performance of EB predictors.

One compares their distribution with estimates obtained through Monte Carlo computations, and the other implements simulations of hypothetical new patients. The model and analysis for the antipsychotic RCT are in Sections 3.1 and 3.2. Results of the evaluations of EB predictors are in Sections 3.3, 3.4 and 3.5. A discussion is in Section 4.

1.2. Methods

1.2.1 Time-dependent personalized medicine models for binary outcomes

We assume that the treatment response is a binary outcome of '1' ('Yes') versus '0' ('No'), with 1 indicating a good condition for the patient. This also applies to controlling the measurement of a continuous or ordinal response such that it is below or above a pre-set value or within a pre-set range, for instance, dichotomizing a response for which the assumptions of alternative regression models may not be valid. For example, as a therapeutic target, we may want to reduce the discrete disorganized dimension score (ranged 0-15) of a psychiatric patient to less than or equal to 3. The response is defined as 1 if the measurement is ≤ 3, and 0 if it is > 4. We also assume that the responses vary over time, which is often the case in medical treatments.

We used mixed-effects logistic regression models to predict the treatment benefits. The binary outcome is denoted as $y_{\omega,j}$ where ω represents a subject (or patient) and j a specific observation at a given time point $t_{\omega,j}$. The 2-PM logistic model is

$$\text{logit}\left(P\left(y_{\omega,j}=1\middle|\boldsymbol{X}_{\omega},t_{\omega,j}\right)\right)=\alpha_{0,\omega}+\boldsymbol{\lambda}^{T}\boldsymbol{X}_{\omega}+\sum_{k=0}^{n}\beta_{k}\,g_{k}(t_{\omega,j})+\sum_{k=1}^{n}\alpha_{k,\omega}\,g_{k}(t_{\omega,j}),\qquad(1)$$

$$\omega=1,\ldots,I,\qquad j=1,\ldots,J_{\omega}$$

where I indicates number of subjects used to estimate model parameters, J indicates number of observations for subject ω, $\boldsymbol{X}_{\omega}$ indicates patient-level covariates (i.e., a subject's characteristics) with fixed effects, g_k are functions of time, β_k and $\boldsymbol{\lambda}$ are fixed effects (population constants). $\alpha_{k,\omega}$ $(k \geq 0)$ are random effects in

the sense that each patient has their own values. Usually, $\alpha_{k,\omega}$ are considered normally distributed with mean 0 (Hedeker & Gibbons, 2006; White et al., 2003). When $\alpha_{k,\omega} = 0$ for $k \geq 1$, Formula (1) reduces to a 1-PM model.

A usual choice for $g_k(t)$ is t^k, $k = 0, \dots, n$, which models the evolution of the response over time with a polynomial trend of degree n. Here, however, $g_0(t), \dots, g_n(t)$ represent orthogonal polynomials of degree $0, \dots, n$, respectively, which facilitate numerical computations and are interpreted similarly (see Section 2.4) (Emerson, 1968; Pettofrezzo, 1984; Hamming, 1987; Hedeker and Gibbons, 2006).

1.2.2 Disease severity and individual benefits

Once model parameters are estimated, we can use them to "predict" (estimate) individual treatment benefits, not only for patients from the clinical trial but also for new patients. The severity of a patient's chronic disease at a given time point is defined as the probability that the patient's response is outside the therapeutic target (Diaz, 2016). At time 0, that is, before treatment starts, the severity is

$$s_0 = 1 - \left(1 + \exp\left(-\alpha_0 - \boldsymbol{\lambda}^T \boldsymbol{X} - \sum_{k=0}^{n} \beta_k \, g_k(0)\right)\right)^{-1}$$

where the index ω is not written in the equation to emphasize that the patient may be a new patient and α_0 is a patient-specific intercept.

The severity for the patient at time t post treatment initiation is

$$s_t = 1 - \left(1 + \exp\left(-\alpha_0 - \boldsymbol{\lambda}^T \boldsymbol{X} - \sum_{k=0}^{n} \beta_k \, g_k(t) - \sum_{k=1}^{n} \alpha_k \, g_k(t)\right)\right)^{-1}$$

The individual benefit of the treatment is the reduction in disease severity from time 0 (Diaz, 2016, 2019). Thus, the patient's benefit after t units of time under treatment is

$$b(t; \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{X}) = s_0 - s_t \qquad (2)$$

where $\boldsymbol{\beta} = (\beta_0, \dots, \beta_n)^T$, and $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_n)^T$ is the vector of patient-specific random effects.

1.2.3 Empirical Bayesian prediction of individual benefits

For calculating the predicted individual treatment benefits at a given time point, we need to estimate the random effects for each patient. For the current study, the command for multilevel mixed-effects generalized linear models ("meglm") in the Stata software was used to fit the mixed-effects logistic model and obtain EB means as predictors of the random effects $\boldsymbol{\alpha}$ (version 15.1, StataCorp LLC, College Station, TX). Once parameter estimates are obtained for the mixed-effects model, Stata's "predict" command only needs the responses and covariates of a patient, either from the original sample or as a new patient, to predict the patient's random effects. The command combines the specific patient's data with the estimated model parameters to compute the predictions. The EB predictor of the patient's random effects is an estimate of the mean of the conditional (posterior) distribution of the random effects given the patient's data. Stata's predict command computes this estimate using adaptive Gaussian quadrature (Skrondal and Rabe-Heketh, 2004).

If $\widehat{\boldsymbol{\alpha}}$ is the EB predictor of a patient's $\boldsymbol{\alpha}$, the EB predictor of the individual benefit at time $t \geq 0$ is

$$b\left(t;\ \hat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{X}\right) \qquad (3)$$

where $\hat{\boldsymbol{\lambda}}$ and $\widehat{\boldsymbol{\beta}}$ are the maximum likelihood estimates of $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$.

Here we adopt standard EB terminology and use the term "predictor" to refer to an estimator of a random coefficient or an individual benefit, which are random variables at the patient population level (Robinson, 1991). In this sense, the term prediction does not refer to the forecast of future values of y. We restrict the term "estimator" to estimators of fixed effects or variance components.

1.2.4 Application to an antipsychotic RCT: retrospective empirical Bayesian prediction of benefits

The antipsychotic-naïve patients with non-affective psychosis provided a written informed consent to be included in the RCT (Pelayo-Teran et al., 2014), which conformed to international standards for research ethics and was approved by the local institutional review board. Here, we analyzed the disorganized dimension scores of the SAPS-SANS scale, with higher scores representing poorer outcomes (Andreasen, 1983a; 1983b). The dichotomous response y was coded as 1 if the subject had a disorganized

dimension score ≤ 3, or 0 otherwise. The responses were available at baseline and at the end of 1, 2, 3, 4, and 6 weeks of antipsychotic treatment. One hundred sixty-one patients were randomized to olanzapine, risperidone or haloperidol (Pelayo-Teran et al., 2014). Since our goal was to measure individual benefits, only the 117 patients with $y = 0$ at baseline were included in these analyses (55 non-cannabis users and 62 cannabis users). Those within the therapeutic target at baseline ($y = 1$) were excluded (26 non-cannabis users and 18 cannabis users).

The final model included cannabis use as a time-independent patient characteristic (Table 1). Therefore, in this application, $\boldsymbol{X}$ included only cannabis use. As in prior analyses (Pelayo-Teran et al., 2014), variable selection for the mixed-effects logistic model did not produce any significant differences among the three antipsychotics. Similarly, diagnosis, duration of untreated psychosis, gender, and smoking did not have significant effects on the odds of being within the therapeutic target. The analyses ruled out the possibility that these variables were confounders of cannabis use and the response.

Table 1. Random intercept logistic regression model of disorganized dimension score less than or equal to 3 from 117 subjects with a first episode of non-affective psychosis under antipsychotic treatment.

Parameter name	Parameter estimate (SE)	p-value	95% CI
Fixed effects			
Cannabis use[a]	-1.647 (0 .7128)	0.021	[-3.044, -0.250]
Orthogonalized time[b,c]	3.722 (0.4242)	<0.0001	[2.890, 4.553]
Orthogonalized time-square[b,d]	-1.911 (0.3079)	<0.0001	[-2.514, -1.307]
Orthogonalized time-cube[b,e]	0.876 (0.2133)	<0.0001	[0.458, 1.294]
Intercept[f]	0.918 (0.5268)	0.081	[-0.115, 1.950]
Variance of random effects			
Intercept	10.681 (3.0136)		[6.144, 18.569]

CI: 95% confidence interval; SE: standard error.
[a]The dichotomous covariate "cannabis use" was defined as 1 if the subject was a cannabis user, 0 otherwise.
[b]Time in weeks was transformed into three mutually orthogonal covariates to build a polynomial of degree 3. The polynomial represented the evolution over time of the logit of the probability of having a disorganized dimension less than or equal to 3.

[c]The 1st order orthogonal polynomial was $g_1(t) = 1.352 + 0.507t$, where t is time. The covariate "orthogonalized time" was computed with this formula.
[d]The 2nd order orthogonal polynomial was $g_2(t) = 1.336 - 1.604t + 0.267t^2$. The covariate "orthogonalized time-square" was computed with this formula.
[e]The 3rd order orthogonal polynomial was $g_3(t) = 1.028 + 3.693t - 1.713t^2 + 0.190t^3$. The covariate "orthogonalized time-cube" was computed with this formula.
[f]The zero-order orthogonal polynomial, $g_0(t)$ is the fixed intercept.

Orthogonal polynomials (Emerson, 1968; Pettofrezzo 1984; Hamming, 1987; Hedeker and Gibbons, 2006) up to degree 3 were used to model the changes of responses over time (**Table 1**). The orthogonal polynomial representation greatly reduces collinearity and scale differences between time powers and simplifies the computation. The transformation of time into orthogonal polynomials is especially useful in mixed-effects models since it speeds up the convergence, which can be challenging for mixed models. No significant random effects for orthogonally-transformed time powers were observed in the random effects logistic model; therefore, only fixed effects were used for the transformed time variables and $\boldsymbol{\alpha}$ included only a random intercept. The Stata command "orthpoly" was used to transform the time variable to orthogonal polynomials (StataCorp LLC, College Station, TX). The "poly" option provided the coefficients of the orthogonal polynomials, allowing treatment benefit prediction at specific time points. The orthogonal polynomials are reported in footnotes c-e in **Table 1**.

Similar to Diaz (2019), we used parameter estimates and data from a specific patient to predict the patient's benefit at time $t + h$. Here, t is the prediction origin, defined as the time up to which the patient's data are collected to make the prediction; and h is the prediction horizon, defined as the elapsed time between the prediction origin and the time for which we want to predict the benefit. For instance, if we have collected data during 3 weeks of treatment and want to predict the patient's benefit at week 5, then $t = 3$ and $h = 2$. If we want to predict the benefit at week 2, then $t = 3$, $h = -1$. If $h \leq 0$, we are retrospectively estimating the benefit achieved at time $t + h$. If $h > 0$, we are forecasting a future benefit value at time $t + h$ (Diaz, 2019). Prediction origin $t = 0$ indicates only baseline responses are available for predicting benefits. Although t and h can be non-integers, we used only integer numbers.

To illustrate how the proposed method can be used in data analysis for retrospective benefit predictions, we predicted the benefits for each of the 55 non-cannabis users and 62 cannabis users at weeks 1 through 6 using Formula (3) and the estimates in Table 1. This formula allows predicting benefits at any given time point, even if the clinical trial did not collect data at that point. To compare the evolutions of individual benefits over time, the sample quartiles of the 62 EB benefit predictions from cannabis users were computed, and similarly for the 55 non-cannabis users (**Table 2**). These sample quartiles can be viewed as estimates of the quartiles of the distributions of individual benefits for the populations of cannabis and non-cannabis users

1.2.5 Assessment of empirical Bayesian predictions

We conducted Monte Carlo computations as an alternative to the EB approach to estimate population quartiles of individual benefits. For cannabis or non-cannabis users at a time point, Monte Carlo estimates of population quartiles were obtained by simulating 1,000 patients assuming the model in **Table 1**, and then calculating the quartiles of their benefits, as described in Supporting Information S1. The estimated population quartiles, reported in **Table 3**, were compared with the sample quartiles of the benefits for the 55 non-cannabis users and 62 cannabis users predicted with the EB approach and reported in **Table 2.** We consider an agreement between these two types of estimates as evidence that EB individual benefit predictors reflect overall population trends reliably.

In addition, a simulation study was conducted to evaluate how well EB benefit prediction would work in new patients. The simulations assessed the performance of Formula (3) for various prediction origins (t), prediction horizons (h), and distances of parameter estimates from true parameters (δ) in standard error units (Diaz, 2017). Spearman's correlations (C_{t+h}) between predicted benefits and simulated true benefits were computed (**Table 4**). Relative biases ($\mathcal{B}_{t+h}$), defined as {(mean of bias) / (mean of true benefit)} x 100 were used to examine prediction accuracy, where bias is the difference between predicted and true benefit (**Table 5**). The simulation methodology is in Supporting Information S2 and S3.

Table 2. Sample medians (and first and third quartiles) of retrospectively-predicted antipsychotic treatment benefits at the end of weeks 1 through 6 for 117 subjects from a pragmatic clinical trial by cannabis use status and time on treatment. Empirical Bayesian predictors of the subject's random effects were used for predicting treatment benefits, combining 6 weeks' data with parameter estimates in Table 1.

Study group	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Non-cannabis (N=55)	0.235 (0.047, 0.828)	0.895 (0.575, 0.894)	0.968 (0.832, 0.989)	0.976 (0.867, 0.990)	0.981 (0.894, 0.990)	0.991 (0.967, 0.994)
Cannabis (N=62)	0.037 (0.003, 0.181)	0.517 (0.069, 0.859)	0.797 (0.214, 0.957)	0.838 (0.264, 0.967)	0.870 (0.316, 0.974)	0.958 (0.614, 0.992)

1.3. Results

1.3.1 The impact of cannabis on individual responses to antipsychotic treatment

A mixed-effects logistic model was used to examine the impact of cannabis use on individual treatment effects of antipsychotics. As described above, the response was defined as 1 if the disorganized dimension score was ≤ 3, or 0 otherwise. The selected 1-PM model had cannabis as well as orthogonalized time, time-square and time-cube as covariates with fixed effects, and a random intercept (**Table 1**). The likelihood of not being in the therapeutic target followed a cubic-polynomial trend over time. On average, cannabis use was significantly associated with decreased odds of having a disorganized dimension ≤ 3, with an odds ratio of 0.193 (95% CI: 0.048 to 0.779) as compared to no use.

1.3.2 Retrospectively predicted antipsychotic benefits

The antipsychotic benefits the subjects received during the RCT were analyzed retrospectively for treatment durations of 1 to 6 weeks. Please note that although the RCT measured patients' responses only at the end of weeks 1, 2, 3, 4, and 6, benefits can be predicted for any treatment duration between 0 and 6 weeks using Formula (3). The medians and the first and third quartiles of the predicted benefits for non-cannabis and cannabis users are shown in **Table 2.**

The quartiles of the benefits for cannabis users were much smaller at earlier weeks as compared to non-cannabis users, indicating generally slower responses to the treatment in cannabis users (**Table 2**). For instance, in non-cannabis users, the median decrease in disease severity was 0.235 probability units compared to 0.037 for cannabis users at week 1. Treatment benefits tended to increase with time for both groups. By weeks 5 or 6, the medians of the benefits are comparable for cannabis users and non-cannabis users; however, the first quartile for cannabis users remained much smaller than that for non-cannabis users, indicating that there were more cannabis users receiving little benefits than non-cannabis users.

1.3.3 Comparison of quartile estimates based on empirical Bayesian predictors with Monte Carlo estimates

The patterns of the evolution of EB predicted benefits (**Table 2**) were like those of the benefit evolution suggested by the Monte Carlo approach (**Table 3)**, indicating a reliable estimation of quartiles of

treatment benefits when using EB predictors of random effects. Both tables reveal a negative impact of cannabis on antipsychotic treatment benefits by delaying the responses in some patients, suggesting a moderating effect for cannabis. Although the number of non-cannabis and cannabis users achieving tangible benefits increased with time, cannabis users achieved benefits more slowly. Even at the end of week 6, there were more cannabis users than non-cannabis users who had not received high benefits yet.

The variations of the EB predicted benefits over time and the differences between cannabis and non-cannabis users are also illustrated using histograms in **Figure 1,** and analogously for the benefits of the 1,000 simulated patients per cannabis status group from Monte Carlo computations shown in **Figure 2**. The two figures exhibit similar patterns, suggesting the adequacy of EB predictors for detecting overall group trends. For each time point, there were more cannabis users who were not receiving substantial benefits compared to non-cannabis users. Even at week 6, there were more cannabis users whose treatment benefits remained minimal.

To visualize how the medians of the retrospectively- predicted EB benefits changed over time for non-cannabis versus cannabis users and compare their patterns with the Monte-Carlo computed medians, we plotted the medians from **Tables 2** and **3** in Panels A and B of **Figure 3**, respectively. The medians of benefits for cannabis users increased at a slower pace compared to non-cannabis users. The patterns for medians of EB predictions (Panel A) are consistent with those for Monte Carlo medians (Panel B), suggesting that the medians of EB retrospective predictions accurately captured group trends in benefit evolution.

1.3.4 Evaluation of EB benefit prediction in simulated new patients using correlations between predictions and true benefits

To examine the performance of the benefit predictor in Formula (3), we analyzed the correlations between the predicted individual benefits and the true individual benefits from simulated new patients using Spearman's correlations (C_{t+h}). Each C_{t+h} was calculated from 1,000 simulated cannabis users or 1,000 non-cannabis users. Results for cannabis users are shown in **Table 4**. See the Supporting Information for non-cannabis users.

Table 3. Estimates of medians (and first and third quartiles) of individual antipsychotic treatment benefits at the end of weeks 1 through 6 in non-cannabis users and cannabis users, obtained with Monte Carlo computation. The model in Table 1 was used for simulating 1,000 patients for each study group.

Study group	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Non-cannabis	0.213 (0.034, 0.665)	0.860 (0.474, 0.962)	0.942 (0.742, 0.983)	0.952 (0.786, 0.986)	0.959 (0.822, 0.988)	0.982 (0.919, 0.994)
Cannabis	0.051 (0.007, 0.324)	0.591 (0.159, 0.914)	0.834 (0.408, 0.970)	0.868 (0.476, 0.975)	0.891 (0.532, 0.979)	0.963 (0.789, 0.991)

Table 4. Spearman correlations (C_{t+h}) between EB predicted benefits and true benefits of antipsychotic treatment in simulated new patients who are cannabis users, by prediction origin (t), prediction horizon (h) and distance of parameter estimates from true parameters in standard error units (δ).

		$t+h$ (weeks)		
t (weeks)	δ	2	4	6
0	0	-0.07	-0.03	0.01
	0.5	-0.07 (-0.17, 0.05)	-0.03 (-0.17, 0.08)	-0.01 (-0.14, 0.15)
	1	-0.07 (-0.31, 0.08)	-0.05 (-0.32, 0.17)	0.01 (-0.28, 0.26)
	1.5	-0.06 (-0.49, 0.16)	-0.06 (-0.49, 0.32)	-0.04 (-0.46, 0.33)
2	0	0.80	0.60	0.61
	0.5	0.79 (0.75, 0.82)	0.59 (0.47, 0.68)	0.62 (0.50, 0.68)
	1	0.77 (0.70, 0.82)	0.60 (0.22, 0.68)	0.62 (0.29, 0.69)
	1.5	0.75 (0.59, 0.80)	0.59 (-0.05, 0.70)	0.59 (0.06, 0.67)
4	0	0.87	0.75	0.68
	0.5	0.87 (0.83, 0.90)	0.77 (0.67, 0.84)	0.67 (0.50, 0.77)
	1	0.87 (0.77, 0.90)	0.77 (0.51, 0.86)	0.69 (0.19, 0.83)
	1.5	0.85 (0.68, 0.90)	0.75 (0.25, 0.86)	0.69 (0.01, 0.84)
6	0	0.89	0.82	0.74
	0.5	0.89 (0.85, 0.92)	0.80 (0.69, 0.87)	0.74 (0.58, 0.83)
	1	0.89 (0.79, 0.93)	0.81 (0.55, 0.89)	0.74 (0.41, 0.87)
	1.5	0.88 (0.71, 0.93)	0.81 (0.27, 0.91)	0.75 (0.16, 0.89)

Note: Parameter estimates in the second column of Table 1 were used for predicting treatment benefits. δ = 0 corresponds to the ideal situation when parameter estimates are equal to the true model parameters, in which case there is only one C_{t+h}. For $\delta > 0$, each entry in the table gives the median (minimum, maximum) of 64 ($=2^6$) values of C_{t+h} corresponding to different combinations of parameter values that are at a distance of δ standard errors from their corresponding estimates. Each value of C_{t+h} was computed using 1,000 simulated new patients who were cannabis users.

Minimal correlations were observed when only baseline data were used for predictions ($t = 0$). When prediction origin $t \geq 1$, relatively high correlations between the predicted benefits and the corresponding true benefits were observed, especially when the parameter estimates were at a distance of 1 standard error or less ($\delta \leq 1$) from their corresponding true parameters (see Supporting Information for results at $t = 1$, 3, and 5). This is most apparent if parameter estimates are the same as the true model parameters ($\delta = 0$). When parameter estimates moved further away from the true parameter values, that is as δ grew, the range of possible correlations grew wider, as expected. However, the median of the correlations stayed approximately the same as for $\delta = 0$. The correlations were relatively high for retrospective predictions not only when predicting the benefits achieved up to the current week ($h = 0$) but also for predictions of past benefits ($h < 0$). Correlations for prospective predictions ($h > 0$) were slightly lower than those for retrospective predictions with comparable t and δ but still above 0.5, indicating a relatively reliable forecasting of future treatment benefits with a prediction horizon of 4 weeks or less, especially when $\delta \leq 1$.

1.3.5 Evaluation of benefit prediction in simulated new patients using relative biases ($\mathcal{B}_{t+h}$)

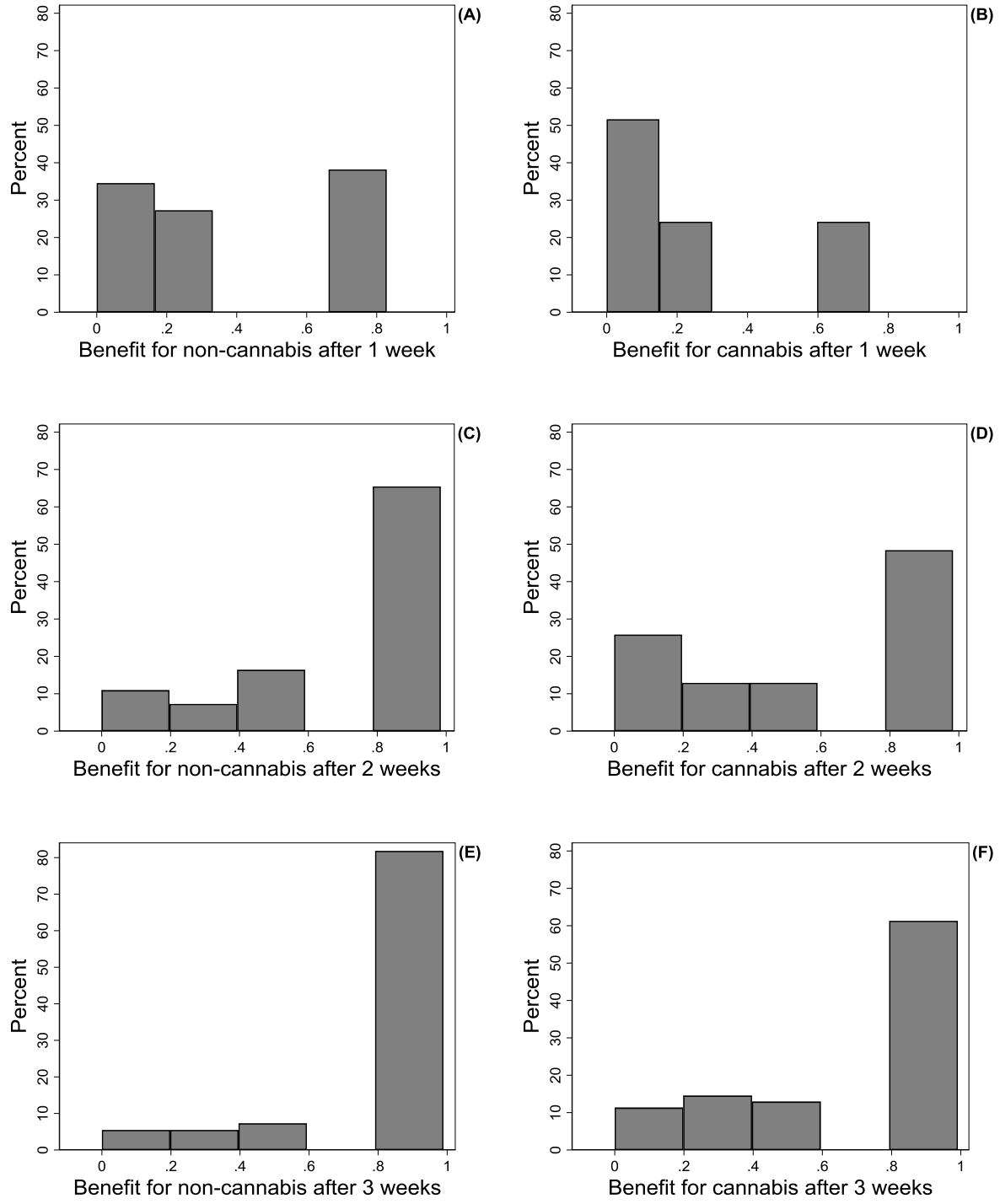
To further examine the performance of Formula (3), we assessed the biases of the predicted benefits relative to the true benefits. Each $\boldsymbol{\mathcal{B}_{t+h}}$ was calculated from 1,000 simulated new cannabis users, and the results are shown in **Table 5.** Negative signs indicate that the predicted benefits are smaller than the true benefits. In general, when predicting the benefits for a given time point, the relative bias $\boldsymbol{\mathcal{B}_{t+h}}$ decreased as the prediction origin (t) increased, indicating that the more data we can use the less biased the prediction. For $t > 0$ and $\delta \leq 1$, the $\boldsymbol{\mathcal{B}_{t+h}}$ were relatively small, suggesting accurate predictions of past, current and future benefits when the patient provides at least one post-treatment response measure, even if the parameter estimates somewhat differ from their corresponding true parameter values. As expected, the range of possible values of $\boldsymbol{\mathcal{B}_{t+h}}$ became wider as δ increased. However, the median of $\boldsymbol{\mathcal{B}_{t+h}}$ stayed approximately the same as for $\delta = 0$.

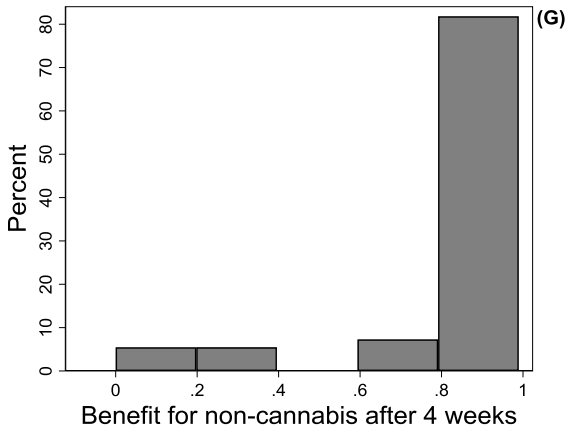
Table 5. Relative biases ($\mathcal{B}_{t+h}$) of empirical Bayesian predictions of antipsychotic treatment benefits in simulated new patients who are cannabis users, by prediction origin (t), prediction horizon (h) and distance of parameter estimates from true parameters in standard error units (δ).

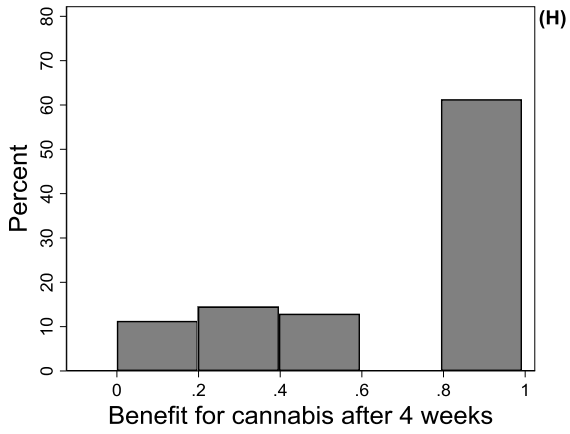
		$t+h$ (weeks)		
t (weeks)	δ	2	4	6
0	0	8.6	24.4	17.6
	0.5	10.1 (-8.1, 36.3)	25.1 (8.3, 48.9)	16.9 (7.2, 33.8)
	1	10.2 (-20.3, 82.1)	25.3 (-0.8, 82.7)	17.4 (1.0, 65.7)
	1.5	9.5 (-29.2, 156.3)	25.9 (-7.5, 139.2)	16.5 (-2.3, 111.4)
2	0	-7.0	-4.2	2.7
	0.5	-5.9 (-10.5, 1.0)	-2.9 (-10.9, 8.7)	3.3 (-4.3, 16.1)
	1	-4.9 (-10.8, 5.8)	-3.8 (-16.4, 25.8)	3.7 (-8.7, 40.8)
	1.5	-4.6 (-12.8, 26.8)	-2.1 (-23.7, 56.4)	3.8 (-13.3, 78.0)
4	0	0.7	-1.3	-1.0
	0.5	-0.5 (-5.4, 7.0)	-1.4 (-5.8, 3.7)	-1.0 (-8.0, 8.2)
	1	-0.4 (-9.6, 18.8)	-1.1 (-8.2, 13.7)	-1.2 (-14.5, 27.1)
	1.5	0.2 (-13.6, 42.6)	-0.9 (-11.9, 35.6)	-1.2 (-22.1, 60.3)
6	0	0.4	-1.0	-1.6
	0.5	0.1 (-7.4, 11.1)	-0.3 (-4.4, 5.7)	-1.2 (-5.4, 6.0)
	1	0.2 (-12.6, 31.7)	-0.4 (-6.4, 12.5)	-1.3 (-9.8, 19.8)
	1.5	-0.7 (-25.4, 63.1)	0.3 (-10.2, 28.0)	-1.1 (-14.3, 46.7)

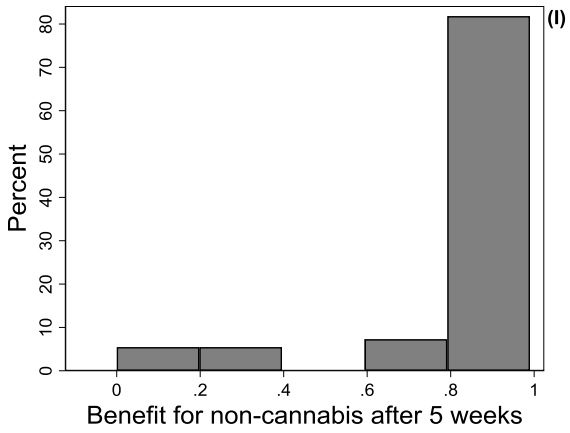
Note: Parameter estimates in the second column of Table 1 were used for predicting treatment benefits. δ = 0 corresponds to the ideal situation when parameter estimates are equal to the true model parameters, in which case there is only one $\mathcal{B}_{t+h}$. For δ > 0, each entry in the table gives the median (minimum, maximum) of 64 (=2^6) values of $\mathcal{B}_{t+h}$ corresponding to different combinations of parameter values that are at a distance of δ standard errors from their corresponding estimates. Each value of $\mathcal{B}_{t+h}$ was computed using 1,000 simulated new patients who were cannabis users.

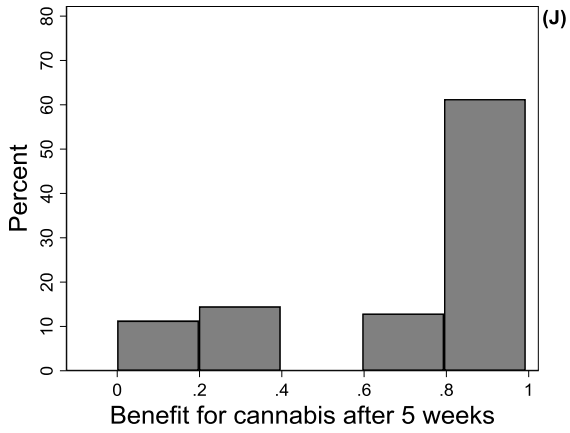
Figure 1. Histograms of retrospectively-predicted antipsychotic treatment benefits at weeks 1 through 6 for the 55 non-cannabis users and 62 cannabis users in the pragmatic clinical trial. Benefits were predicted with the empirical Bayesian approach [Formula (3)]. (A), (C), (E), (G), (I) and (K) are predicted benefits for non-cannabis users at weeks 1, 2, 3, 4, 5 and 6, respectively; (B), (D), (F), (H), (J) and (L) are predicted benefits for cannabis users at weeks 1, 2, 3, 4, 5 and 6, respectively.

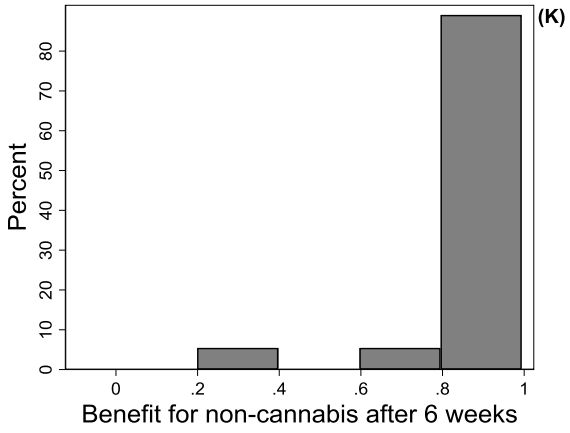



(G)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 4 weeks


(H)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for cannabis after 4 weeks


(I)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 5 weeks


(J)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for cannabis after 5 weeks


(K)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 6 weeks

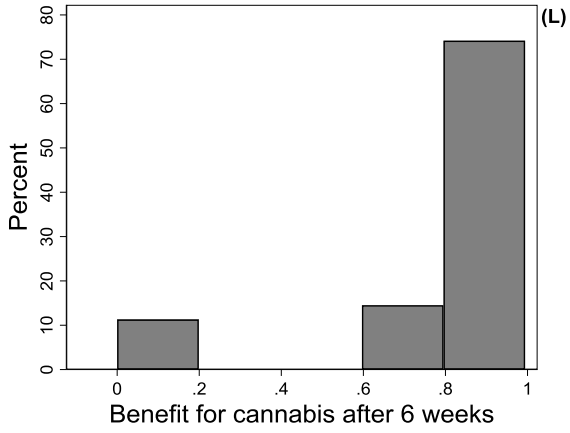
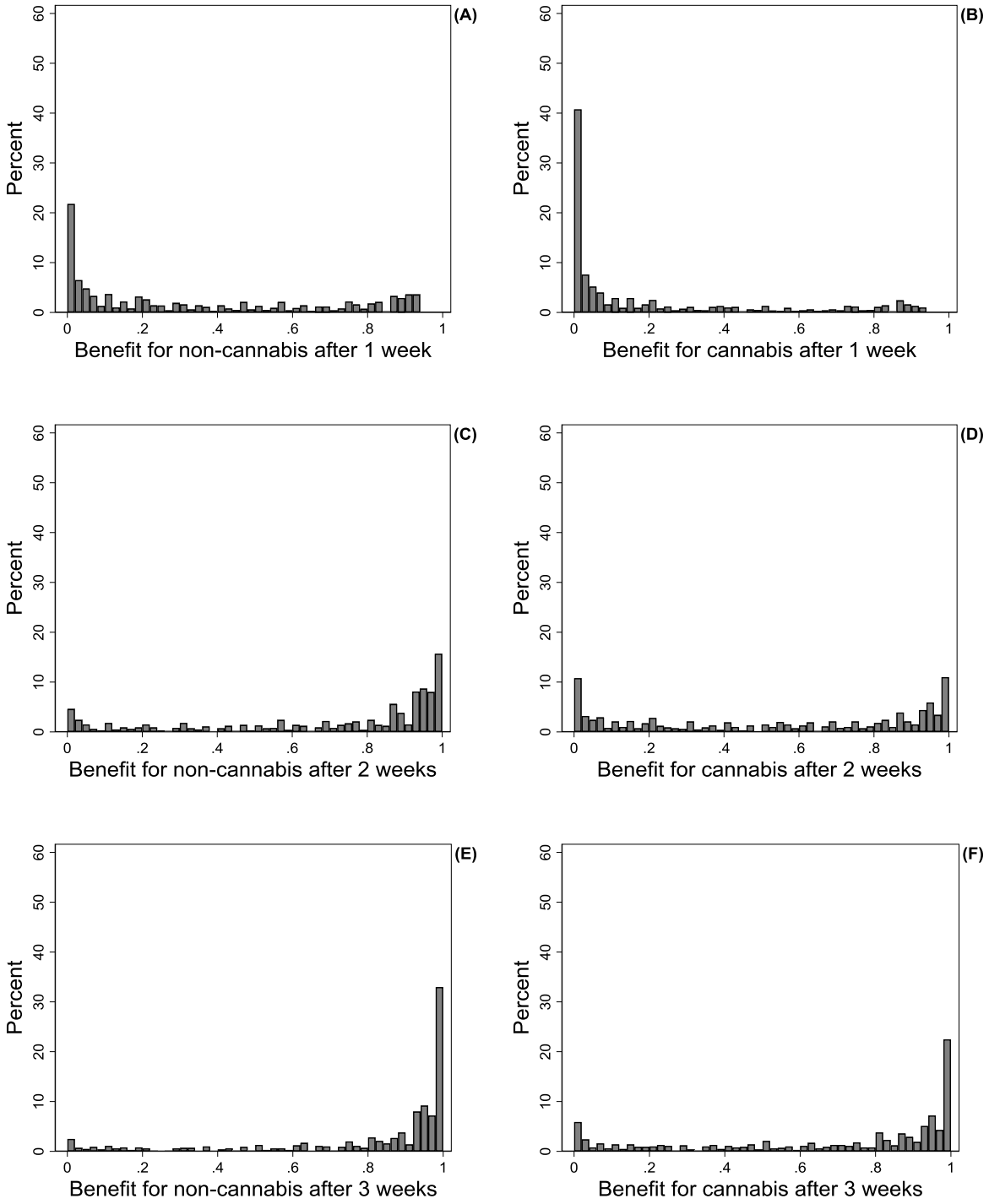
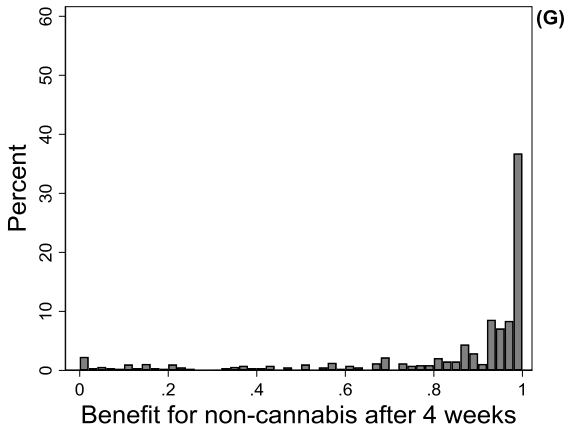
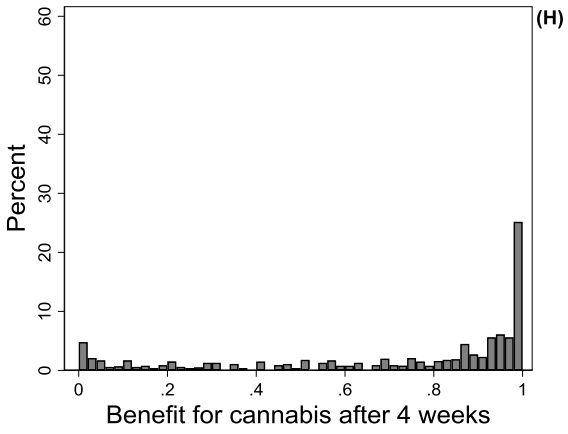
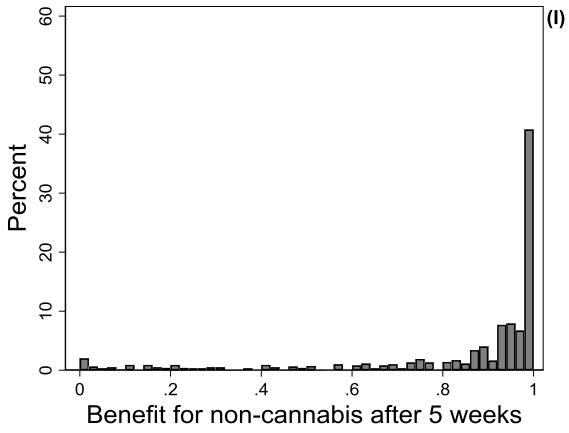

(L)
Percent
0
10
20
30
40
50
60
70
80
0
.2
.4
.6
.8
1
Benefit for cannabis after 6 weeks

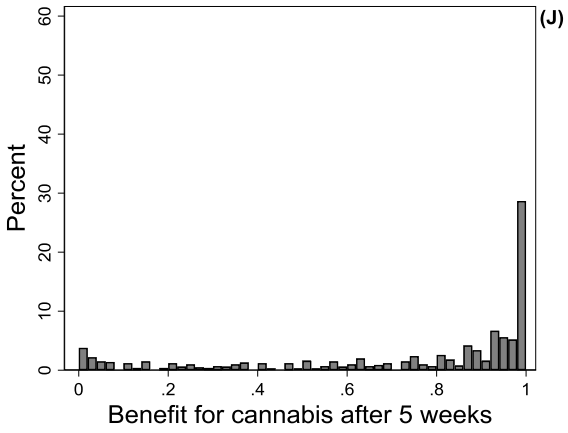
Figure 2. Histograms of individual antipsychotic treatment benefits at weeks 1 through 6 from 1,000 simulated non-cannabis users and 1,000 simulated cannabis users, assuming the model in Table 1. (A), (C), (E), (G), (I), and (K) are benefits for non-cannabis users at weeks 1, 2, 3, 4, 5, and 6, respectively; (B), (D), (F), (H), (J), and (L) are benefits for cannabis users at weeks 1, 2, 3, 4, 5, and 6, respectively.

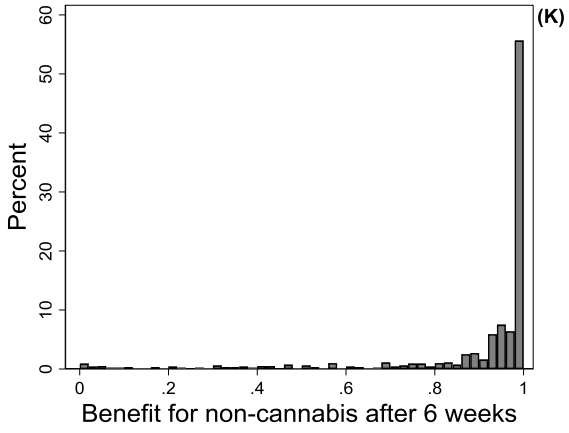



(G)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 4 weeks


(H)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for cannabis after 4 weeks


(I)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 5 weeks


(J)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for cannabis after 5 weeks


(K)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for non-cannabis after 6 weeks

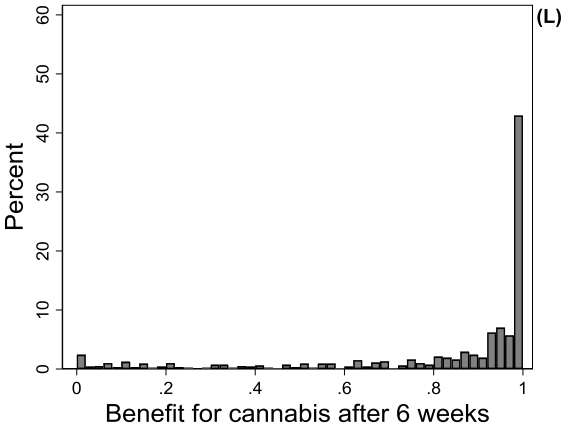
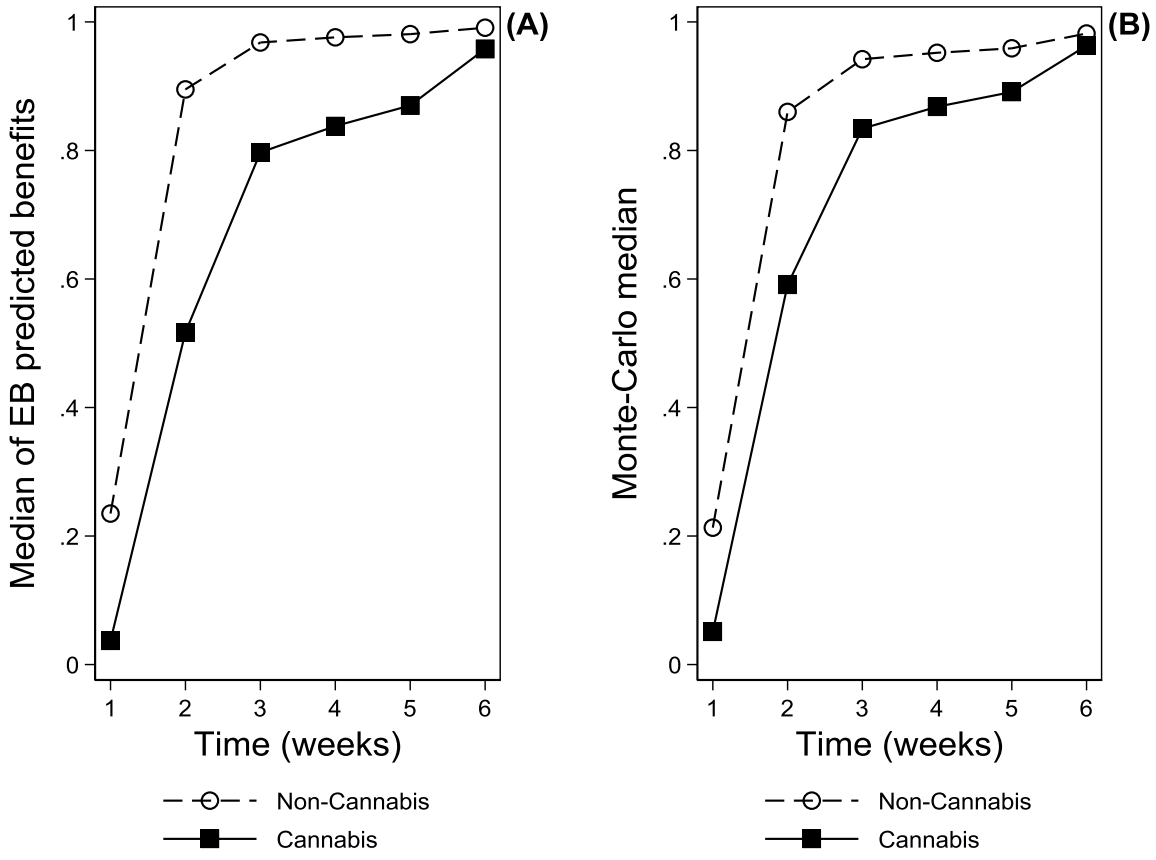

(L)
Percent
0
10
20
30
40
50
60
0
.2
.4
.6
.8
1
Benefit for cannabis after 6 weeks

Figure 3. Comparison of estimators of medians of individual antipsychotic benefits at weeks 1 through 6. (A) Plots of medians of retrospectively predicted antipsychotic treatment benefits for the 55 non-cannabis users and 62 cannabis users from the pragmatic clinical trial, using the empirical Bayesian approach. (B) Medians of the individual benefits of 1,000 non-cannabis users and 1,000 cannabis users were simulated assuming the Model in Table 1.



1.4. Discussion

In this paper, we evaluated EB predictors of individual treatment benefits in the context of longitudinal binary outcomes which are frequent in medical research. Our results suggest that EB predictors accurately capture overall population trends in the achievement of individual benefits and show that EB predictors will reliably measure individual benefits in new patients both retrospectively and prospectively.

Our approach utilizes EB predictors of individual random effects that are plugged into the formula defining benefit functions (Formula 3). The method is applicable to both 1-PM and 2-PM models (Diaz,

2016, 2019) and can be used to retrospectively evaluate patients' responses from a RCT. It can also be used to retrospectively or prospectively predict individual benefits of different treatments in new patients with known characteristics and previous responses. Standard statistical packages implementing mixed-effects logistic models such as Stata (StataCorp LLC, College Station, TX) or SAS (SAS Institute Inc., Cary, NC) can be used to compute EB predictors.

As an application, we used data from an antipsychotic RCT in patients with a first episode of non-affective psychosis (Pelayo-Teran et al., 2014) and fitted a 1-PM model for the dichotomized disorganized dimension. Simulations showed that EB prediction of benefits was reliable, with small relative biases and relatively high correlations between predicted and true benefits, except when only baseline data are available for predictions.

The present study confirmed cannabis use as a negative prognostic predictor for the disorganized dimension during antipsychotic treatment. Cannabis users were found to respond less and more slowly as compared to non-cannabis users based on individual benefit measurements, which is consistent with an analysis of the same data using a censored normal model of response trajectories that quantified only average cannabis effects (Pelayo-Teran et al., 2014). Our results support earlier findings that cannabis use is associated with more severe psychotic symptoms in patients with psychosis (Caspari, 1999; Foti et al., 2010; Grech et al., 2005; Kuepper et al., 2011; Zammit et al., 2008).

There is strong experimental evidence that cannabis use may cause psychotic symptoms (Bhattacharryya et al., 2009, 2012, 2015). In a carefully designed cross-over study, Bhattacharryya et al. (2015) randomized 36 healthy subjects to either the sequence of 10 mg of delta-9-tetrahydrocannabinol (delta-9-THC) and placebo or vice versa. The subjects did not have a personal or family history of mental illness, had minimal use of cannabis, alcohol or other psychotropic drugs and refrained from consuming caffeine, alcohol or tobacco during the study. Relative to placebo, the acute administration of delta-9-THC significantly induced the appearance of psychotic-like symptoms and anxiety (Bhattacharryya et al., 2015). This suggests that the reported associations between cannabis use and psychosis severity from observational

studies, or from experiments without randomization to cannabis or non-cannabis use, are not just the result of uncontrolled confounding factors.

It is noteworthy that the 1-PM model with just a random intercept (and no interaction terms between cannabis and time or random effect for time) can show how the benefits evolve differently over the treatment period depending on patient characteristics, in this case cannabis use. This supports Diaz's observation that individual benefit prediction reveals aspects of clinical phenomena that regression models alone cannot show (Diaz 2016, 2019). In this sense, it is a useful complement to standard regression analyses. In fact, we did not find any significant interaction between cannabis use and time when following the standard approach of testing the significance of the product of these two variables (data not shown). We were able to show, however, that cannabis use *modified* the effect of antipsychotic treatment in a time-dependent way.

At week 6 post-treatment, although the median individual benefits for cannabis users was comparable to that for non-cannabis users (**Tables 2 and 3**), there were more cannabis users without substantial treatment benefits (**Figures 1** and **2**). EB prediction makes it feasible to visualize the variation of treatment benefits among patients with different known characteristics as well as with the same known characteristics.

We used correlations and relative biases to evaluate how well EB-predicted individual benefits measure true benefits (Tables 4 and 5). The correlations and relative biases were poor when only baseline data were available, which makes sense since the model did not include a random effect for time that was correlated with the intercept. Various δ values (0, 0.5, 1, and 1.5) were used to mimic the fact that the true parameter values may differ from the parameter estimates. As expected, the ranges for both correlations and relative biases became wider as δ increased (the less precise the estimators were, the less reliable predictions were). Nevertheless, the medians of correlations and relative biases remained stable, suggesting some robustness of EB benefit predictors to imprecise parameter estimation.

The proposed method offers an excellent tool for analyzing clinical trials with binary outcomes that evolve over time. In the example application there were no variables in the model representing the three

different antipsychotics used in the trial because their effects did not significantly differ. The method, however, does allow inclusion of covariates representing treatment options. As such, it can be used to compare individual benefits for various treatments and help clinicians choose medications with the most promising benefits for new patients using patients' characteristics and previous responses. Moreover, we can utilize existing software for computing EB predictors of random effects to predict individual benefits. This makes the application of this method more practical in clinical trial data analysis and potentially in medical practice.

1.5. Limitations

Our data do not allow completely establishing a causal relationship between cannabis and response to antipsychotics, because subjects were not randomized to cannabis or non-cannabis use. However, our results are consistent with other studies that show an association between cannabis use and increased risk of developing psychotic disorders as well as adverse outcomes in patients with psychosis (Linszen et al., 1994; Moore et al., 2007; Zammit et al., 2008), and are consistent with experimental evidence of causality in this association (Bhattacharryya et al., 2009, 2012, 2015).

In the application, we excluded 44 patients (26 non-cannabis and 18 cannabis users) whose disorganized dimension score at baseline was within the treatment target (≤ 3). More non-cannabis users were excluded, as expected, which could have potentially biased the regression results in favor of the null hypothesis of no difference between the two groups. Thus, if such bias occurred, the differences in antipsychotic benefits between the two populations may be greater than the differences observed in our patient sample. This, however, would not invalidate the model as a predictor of individual benefits in new patients because the model would be applicable only to patients with scores outside the treatment target.

1.6. References

Andreasen, N.C. (1983a). *Scale for the Asssessment of Negative Symptoms (SANS)*. Iowa City: University of Iowa.

Andreasen, N.C. (1983b). *Scale for the Asssessment of Positive Symptoms (SAPS).* Iowa City: University of Iowa.

Andrews, N., Cho, H. (2018). Validating effectiveness of subgroup identification for longitudinal data. *Statistics in Medicine* 37:98-106.

Barrigón, M.L., Gurpegui, M., Ruiz-Veguilla, M., Diaz, F.J., Anguita, M., Sarramea, F., Cervilla, J. (2010) Temporal relationship of first-episode non-affective psychosis with cannabis use: A clinical verification of an epidemiological hypothesis. *Journal of Psychiatric Research* 44:413-420.

Bhattacharyya, S., Fusar-Poli, P., Borgwardt, S., Martin-Santos, R., Nosarti, C., O'Carroll, C., Allen, P., Seal, M.L., Fletcher, P.C., Crippa, J.A., Giampietro, V., Mechelli, A., Atakan, Z., McGuire, P. (2009). Modulation of mediotemporal and ventrostriatal function in humans by Delta9-tetrahydrocannabinol: a neural basis for the effects of Cannabis sativa on learning and psychosis. *Archives of Genaral Psychiatry* 66:442-451.

Bhattacharyya, S., Atakan, Z., Martin-Santos, R., Crippa, J.A., Kambeitz, J., Prata, D., Williams, S., Brammer, M., Collier, D.A., McGuire, P.K. (2012) Preliminary report of biological basis of sensitivity to the effects of cannabis on psychosis: AKT1 and DAT1 genotype modulates the effects of δ-9-tetrahydrocannabinol on midbrain and striatal function. *Molecular Psychiatry* 17:1152-1155.

Bhattacharyya, S., Atakan, Z., Martin-Santos, R., Crippa, J.A., Kambeitz, J., Malhi, S., Giampietro, V., Williams, S., Brammer, M., Rubia, K., Collier, D.A., McGuire, P.K. (2015). Impairment of

inhibitory control processing related to acute psychotomimetic effects of cannabis. *European Neuropsychopharmacology* 25:26-37.

Boulesteix, A.L., De Bin, R., Jiang, X., Fuchs, M. (2017). IPF-LASSO: Integrative L(1)-penalized regression with penalty factors for prediction based on multi-omics data. *Computational and Mathematical Methods in Medicine* 2017:7691937.

Botts, S., Diaz, F. J., Santoro, V., Spina, E., Muscatello, M. R., Cogollo, M., Castro, F. E., de Leon, J. (2008). Estimating the effects of co-medications on plasma olanzapine concentrations by using a mixed model. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 32:1453-1458.

Caspari, D. (1999) Cannabis and schizophrenia: results of a follow-up study. *European Archives of Psychiatry and Clinical Neurosciences* 249:45-49.

Cho, H., Wang, P., Qu, A. (2017). Personalized treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica* 27:187-205.

Clausen, L., Hjorthøj, C.R., Thorup, A., Jeppesen, P., Petersen, L., Bertelsen, M., Nordentoft M. (2014). Change in cannabis use, clinical symptoms and social functioning among patients with first-episode psychosis: a 5-year follow-up study of patients in the OPUS trial. *Psychological Medicine* 44:117-126.

Cobo, B., Rueda, M.M., López-Torrecillas, F., (2017). Application of randomized response techniques for investigating cannabis use by Spanish university students. *International Journal of Methods in Psychiatric Research* 26(4). doi: 10.1002/mpr.1517.

Dekker, N., Koeter, M., Den Brink, W. GROUP Investigators. (2012) Craving for cannabis in patients with psychotic disorder, their non-affected siblings and healthy controls: psychometric analysis of the obsessive-compulsive drug use scale. *International Journal of Methods in Psychiatric Research* 21:286-300.

de Leon, J. (2012). Evidence-based medicine versus personalized medicine: are they enemies? *Journal of Clinical Psychopharmacology* 32:153-164.

Diaz, F.J. (2016). Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. *Statistics in Medicine* 35:4077-4092.

Diaz, F.J. (2017). Estimating individual benefits of medical or behavioral treatments in severely ill patients. *Statistical Methods in Medical Research* 28:911-927.

Diaz, F.J., Berg, M.J., Krebill, R., Welty, T., Gidal, B.E., Alloway, R., Privitera, M. (2013b). Random-effects linear modeling and sample size tables for two special crossover designs of average bioequivalence studies: the four-period, two-sequence, two-formulation and six-period, three-sequence, three-formulation designs. *Clinical Pharmacokinetics* 52:1033-1043.

Diaz, F.J., Cogollo, M.R., Spina, E., Santoro, V., Rendon, D.M., de Leon, J. (2012b). Drug dosage individualization based on a random-effects linear model. *Journal of Biopharmaceutical Statistics* 22:463-484.

Diaz, F.J., de Leon, J. (2013a) The mathematics of drug dose individualization should be built with random effects linear models. *Therapeutic Drug Monitoring* 35:276-277.

Diaz, F.J., Eap, C.B., Ansermot, N., Crettol, S., Spina, E., de Leon, J. (2014) Can valproic acid be an inducer of clozapine metabolism? *Pharmacopsychiatry* 47:89-96.

Diaz, F.J., Rivera,T.E., Josiassen, R.C., de Leon, J. (2007). Individualizing drug dosage by using a random intercept linear model. *Statistics in Medicine* 26:2052-2073.

Diaz, F.J., Santoro, V., Spina, E., Cogollo, M., Rivera, T.E., Botts, S., de Leon, J. (2008) Estimating the size of the effects of co-medications on plasma clozapine concentrations using a model that controls for clozapine doses and confounding variables. *Pharmacopsychiatry* 41:81-91.

Diaz, F.J., Yeh, H-W., de Leon, J. (2012a) Role of statistical random-effects linear models in personalized medicine. *Current Pharmacogenomics and Personalized Medicine* 10:22–32.

Emerson, P.L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics* 24:695-701.

Foti, D.J., Kotov, R., Guey, L.T., Bromet, E.J. (2010). Cannabis use and the course of schizophrenia: 10-year follow-up after first hospitalization. *American Journal of Psychiatry* 167:987-993.

Goldstein, B.A., Navar, A.M., Carter, R.E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European Heart Journal* 38:1805-1814.

Grech, A., Van Os, J., Jones, P.B., Lewis, S.W., Murray, R.M. (2005). Cannabis use and outcome of recent onset psychosis. *European Psychiatry* 20:349-353.

Hamming R. (1987). *Numerical Methods for Scientists and Engineers (2nd ed).* Mineola, NY: Dover.

Hedeker, D., Gibbons, R.D. (2006). *Wiley Series in Probability and Statistics. Longitudinal Data Analysis.* Hoboken, NJ: Wiley-Interscience.

Hides, L., Dawe, S., Kavanagh, D.J., Young, R.M. (2006). Psychotic symptom and cannabis relapse in recent-onset psychosis. Prospective study. *British Journal of Psychiatry* 189:137-143.

Kim, S., Baladandayuthapani, V., Lee, J.J. (2017). Prediction-oriented marker selection (PROMISE): with application to high-dimensional regression. *Statistics in Biosciences* 9:217-245.

Kuepper, R, van Os, J., Lieb, R., Wittchen, H.U., Höfler, M., Henquet, C. (2011). Continued cannabis use and risk of incidence and persistence of psychotic symptoms: 10 year follow-up cohort study. *British Medical Journal* 342:d738.

Legleye, S. (2018). The cannabis abuse screening test and the DSM-5 in the general population: optimal thresholds and underlying common structure using multiple factor analysis. *International Journal of Methods in Psychiatric Research* 27:e1597.

Linszen, D.H., Dingemans, P.M., Lenior, M.E. (1994). Cannabis abuse and the course of recent onset schizophrenic disorders. *Archives of General Psychiatry* 51:273-279.

Linszen, D.H., Dingemans, P.M., Nugter, M.A., Van der Does, A.J, Scholte, W.F., Lenior, M.A. (1997). Patient attributes and expressed emotion as risk factors for psychotic relapse. *Schizophrenia Bulletin* 23:119-130.

Ma, J., Stingo, F.C., Hobbs, BP. (2016). Bayesian predictive modeling for genomic based personalized treatment selection. *Biometrics* 72: 575-583.

Moore, T.H., Zammit, S., Lingford-Hughes, A., Barnes, T.R., Jones, P.B., Burke, M., Lewis, G. (2007). Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet* 370:319-328.

Pelayo-Terán, J.M., Diaz, F.J., Pérez-Iglesias, R., Suárez-Pinilla, P., Tabarés-Seisdedos, R., de Leon, J, Crespo-Facorro, B. (2014). Trajectories of symptom dimensions in short-term response to antipsychotic treatment in patients with a first episode of non-affective psychosis. *Psychological Medicine* 44:37-50.

Pettofrezzo, A.J. (1984). *Introductory Numerical Analysis.* Mineola, NY: Dover.

Powers, S., Qian, J., Jung, K., Schuler, A., Shah, N.H., Hastie, T., Tibshirani, R. (2018) Some methods for heterogeneous treatment effect estimation in high dimensions. *Statistics in Medicine* 37:1767-1787.

Robinson, G.K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects: Rejoinder. *Statistical Science* 6:48-51.

Ruberg, S.J., Chen, L., Wang, Y. (2010). The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. *Clinical Trials* 7:574-583.

Schoeler, T., Kambeitz, J., Behlke, I., Murray, R., Bhattacharyya, S. (2016). The effects of cannabis on memory function in users with and without a psychotic disorder: findings from a combined meta-analysis. *Psychological Medicine* 46:177-188.

Senn, S. (2016). Mastering variation: variance components and personalised medicine. *Statistics in Medicine* 35:966-977.

Sies, A., Demyttenaere, K., Van Mechelen, I. (2019) Studying treatment-effect heterogeneity in precision medicine through induced subgroups. *Journal of Biopharmaceutical Statistics* 29:491-507.

Skrondal, A., Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models (1st edition).* Boca Raton, FL: Chapman and Hall/CRC.

White, I.R., Moodie, E., Thompson, S.G., Croudace, T. A modelling strategy for the analysis of clinical trials with partly missing longitudinal data. *International Journal of Methods in Psychiatric Research* 12:139-150.

Xu ,W., Hedeker, D. (2001). A random-effects mixture model for classifying treatment response in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics* 11:253-73.

Zammit, S., Moore, T.H., Lingford-Hughes, A., Barnes, T.R., Jones, P.B., Burke, M., Lewis, G. (2008). Effects of cannabis use on outcomes of psychotic disorders: systematic review. *British Journal of Psychiatry* 193:357-363.

Zhu, X., Qu, A. (2016). Individualizing drug dosage with longitudinal data. *Statistics in Medicine* 5:4474-4488.

Chapter 2: Predicting individual benefits of medical treatments using longitudinal hospital data with non-ignorable missing responses caused by patient discharge

In collaboration with Drs. Nikos Pantazis, Ph.D. and Francisco J. Diaz, Ph.D.

(See collaborator affiliations in Acknowledgements)

2.0. Abstract

We present a method for predicting individual treatment benefits with a novel 2-PM model that handles non-ignorable missingness due to hospital discharge and evaluate its reliability and accuracy by simulations. The longitudinal outcome of interest is modeled simultaneously with the hospital length of stay. The method was illustrated with an application assessing individual pain management benefits post spine fusion surgery, and the pain scores were pre-transformed with a discrete logit transformation. Empirical Bayes (EB) prediction was used to estimate patient level random effects. EB-Predicted individual benefits were compared with the Monte-Carlo computed benefits. To assess the prediction accuracy, we calculated Pearson's correlation between the predicted and the true benefits as well as relative biases of the predicted benefits. Results showed that the EB-predicted individual benefits are close to Monte-Carlo computed ones. The prediction is reliable given that the parameter estimates are not far from the true parameter values. In summary, we proposed to use a 2-PM model with joint mixed effects to predict individual treatment benefits using unbalanced EHR data. This method will help to gain insights on treatment effects from real-world data.

Keywords: individual benefits, Empirical Bayesian prediction, non-ignorable missingness, random effects, observational data

2.1. Introduction

It is increasingly recognized that a patient's response to a medical treatment is a statistically heterogeneous phenomenon (de Leon, 2012). The average treatment effects may not represent a heterogeneous population of patients (Ruberg et al., 2010). The benefits each patient receive from the treatment could differ, requiring measurement of treatment benefits at the patient level (Diaz, 2016, 2019). Generalized linear mixed-effects models (Andrews and Cho, 2018; Botts et al., 2008; Cho 2017; Diaz, 2016, 2019; Diaz et al., 2007, 2008, 2012a, 2012b, 2013a, 2013b, 2014; Senn, 2016; Zhu and Qu, 2016) allow identifying the various sources of variation of patients' responses (Gewandter et al., 2019), offering an excellent tool for analyzing individual benefits. Diaz (2016, 2019) used 1-dimensional personalized medicine (1-PM) and 2-dimensional personalized medicine (2-PM) models to assess individual treatment benefits for clinical trial data using empirical Bayes (EB) predictors. The EB predictors of individual benefits are obtained using the EB predictor of the patient's random effects as well as the estimated fixed effects. The EB predictor of the random effects is an estimate of the mean of the conditional distribution of the random effects given the patient's data.

It is also increasingly recognized that real-world data (RWD) such as electronic health records (EHR) collected in a non-randomized setting hold critical value for clinical evidence generation and play a complementary role to clinical trial data (Miksad and Abernethy 2018). EHR data provide contextual details and longitudinal follow-up for patient's outcomes. One limitation of the EHR data, however, is that there is usually incomplete follow-up due to hospital discharge. Since hospital discharge often depends on patient responses, the missing responses after discharge are nonignorable missing data (Little and Rubin 2002; Pantazis et al. 2010). This creates a problem for predicting treatment benefits because generalized mixed effects models assume missing at random (Hedeker and Gibbons, 2006; Laird 1998). When the missingness is non-ignorable, the analysis results can be seriously biased (Touloumi et al. 1999).

Here, we propose to measure individual treatment benefits with hospital data by jointly modeling the patients' responses to the medical treatment and hospital length of stay (LOS). Joint mixed-effects

models combining a generalized linear mixed effects model and a survival model have been used to handle longitudinal clinical trial data with informative drop-outs which produce non-ignorable missings (Schluchter 1992; De Gruttola and Tu 1994; Touloumi et al. 1999; Pantazis et al. 2010; Crowther et al. 2012; Armero et al. 2018; Hickey et al. 2018; Shardell and Ferrucci 2018; Schluchter and Piccorelli 2019; Papageorgiou et al. 2019). For example, Touloumi et al. (1999) developed a method of parameter estimation for joint models that combines restricted iterative generalized least-squares with a nested expectation-maximization algorithm. To our knowledge, these models have not been used to model hospital data, which are unavoidably biased by non-ignorable missingness due to discharge.

This study was motivated by the fact that many outcomes of clinical procedures, pharmacological therapies, or patient-reported outcomes measurements recorded in longitudinal EHR data are associated with hospital LOS. For instance, laboratory results such as biological markers of acute myocardial infarction (Gronski et al. 2012) or acute kidney injury (Edelstein 2008), as well as physical/behavioral scores (Shaw et al. 2018), are often measured only during hospital stay and are used in discharge planning and decision making. One example of patients' self-reported measurements is pain scores post a surgical procedure, which are available before surgery or during the hospital stay after the surgery but are no longer recorded after discharge.

This study has three objectives. The first is to extend the methodology for predicting individual benefits in clinical trials (Diaz 2016, 2019) to predicting individual benefits using hospital data with non-ignorable missingness. The second is to extend the definition of 2-PM models to joint mixed effects models that simultaneously represent the longitudinal patients' outcome and the hospital LOS. The third is to evaluate the performance of the EB predictors of individual benefits based on joint mixed models using Pearson's correlations between the predicted and the true benefits and the relative biases of the predicted benefits.

2.2. Methods

2.2.1. Joint model for observational longitudinal continuous outcomes with non-ignorable missingness

Next we describe a joint multivariate random effects model to jointly model a continuous outcome and the hospital LOS. Suppose subject i provided n_i outcome measurements on days $t_1 < \cdots < t_{n_i}$ counted from treatment day $t_1 = 0$. Let $\boldsymbol{y}_i^* = \left(y_{i1}, \ldots, y_{in_i}\right)^T$ be the outcome measurements, where y_{i1} is assumed to be measured before treatment and y_{ik} , $k \geq 2$ are measured after treatment. Let $\boldsymbol{x}_{ij} = \left(x_{ij,1}, \ldots, x_{ij,p}\right)^T$ and $\boldsymbol{z}_{ij} = \left(z_{ij,1}, \ldots, z_{ij,q}\right)^T$ be vectors of covariates obtained at time t_j. A covariate can be time-independent (for instance, gender, race, etc.) or a known function of time (for instance, t, t^2, etc.) The covariates in $\boldsymbol{z}_{ij}$ are usually a subset of the covariates in $\boldsymbol{x}_{ij}$. For subject i, the design matrix for the fixed and random effects of the outcome model are $\boldsymbol{X}_i^* = \left(\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,n_i}\right)^T$ and $\boldsymbol{Z}_i^* = \left(\boldsymbol{z}_{i,1}, \ldots, \boldsymbol{z}_{i,n_i}\right)^T$, respectively. The outcome model is

$$\boldsymbol{y}_i^* = \boldsymbol{X}_i^* \boldsymbol{\beta} + \boldsymbol{Z}_i^* \, \boldsymbol{\alpha}_i + \boldsymbol{e}_i$$

where $\boldsymbol{y}_i^* = \left(y_{i1}, \ldots, y_{in_i}\right)^T$ is a vector containing the outcomes for subject i in time order, $\boldsymbol{\beta}$ is the vector of fixed regression coefficients, $\boldsymbol{\alpha}_i$ is the normally distributed vector of random effects with mean 0, and $\boldsymbol{e}_i$ is the vector of residuals for subject i that are assumed to be independent between subjects and normally distributed with mean 0 and variance-covariance $\boldsymbol{R}_i^* = \sigma_e^2 I_{n_i}$.

Let T_i^d be the hospital LOS in days. We assume that discharge always occurs after the last available outcome measurement, that is, $t_{n_i} < T_i^d$. Thus, if T_i^d was available in the EHR dataset and $t_{n_i} = T_i^d$ we add a small offset (i.e. 0.01 days) to make discharge time slightly larger than the last outcome measurement time. The discharge time is considered censored at $t_{n_i} + 0.01$ if either T_i^d is missing in the dataset or if T_i^d is available but $t_{n_i} \leq T_i^d - 1$.

Let $\boldsymbol{x}_i^d = \left(1, x_{i1}^d, \ldots, x_{ir}^d\right)^T$ be time-independent patient's characteristics possibly related to LOS. The discharge time model is

$$\log\left(T_i^d\right) = \boldsymbol{x}_i^{d\,T} \boldsymbol{\beta}^d + e_i^d \,,$$

where $\boldsymbol{\beta}^d$ is a vector of fixed regression coefficients and $e_i^d \sim N(0, \sigma_d^2)$ is a residual.

The joint multivariate random effects model is

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta}^j + \boldsymbol{Z_i}\,\boldsymbol{\alpha}_i^j + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{y_i} = \begin{bmatrix} \boldsymbol{y}_i^* \\ T_i^d \end{bmatrix}$, $\boldsymbol{X_i} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}_i^* \\ \boldsymbol{x}_i^{d\,T} & \boldsymbol{0}^T \end{bmatrix}$, $\boldsymbol{\beta}^j = \begin{bmatrix} \boldsymbol{\beta}^d \\ \boldsymbol{\beta} \end{bmatrix}$, $\boldsymbol{Z_i} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{Z}_i^* \\ 1 & \boldsymbol{0}^T \end{pmatrix}$, $\boldsymbol{\alpha}_i^j = \begin{bmatrix} e_i^d \\ \boldsymbol{b}_i \end{bmatrix}$, and $\boldsymbol{\varepsilon}_i = \begin{bmatrix} \boldsymbol{e}_i \\ 0 \end{bmatrix}$.

2.2.2. EB prediction of the random effects

The EB predictor of the random effects $\boldsymbol{\alpha}_i^j$ is

$$\boldsymbol{\alpha}_{EB,i}^j = \widehat{\boldsymbol{D}}\boldsymbol{Z_i}^{\boldsymbol{T}}\widehat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{e}}_{\boldsymbol{i}}\,,$$

where $\widehat{\boldsymbol{D}}$ is the estimator of $\boldsymbol{D} = \mathrm{Var}\left(\boldsymbol{b}_i^j\right)$, and $\widehat{\boldsymbol{V}_i}$ is the estimator of $\boldsymbol{V_i} = \mathrm{Var}(\boldsymbol{y}_i) = \boldsymbol{R_i} + \boldsymbol{Z_i}\boldsymbol{D}\boldsymbol{Z_i^T}$ with $\boldsymbol{R_i} = \mathrm{Var}(\,\boldsymbol{\varepsilon}_i) = \begin{pmatrix} \boldsymbol{R}_i^* & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{pmatrix}$, and $\hat{\boldsymbol{e}}_i = \begin{bmatrix} \boldsymbol{y}_i^* - \boldsymbol{X}_i^*\widehat{\boldsymbol{\beta}} \\ 0 \end{bmatrix}$ is the estimated residuals for subject i.

The last row of $\hat{\boldsymbol{e}}_i$ is set to 0 during the calculation of the random effects because the error term of the LOS model (e_i^d) is already included in $\boldsymbol{\alpha}_i^j$.

The 1st element of $\boldsymbol{\alpha}_{EB,i}^j$ is the EB estimate of the LOS model residual for subject i. The other elements of $\boldsymbol{\alpha}_{EB,i}^j$ estimate $\boldsymbol{\alpha}_i$ and are denoted here by $\widehat{\boldsymbol{\alpha}}_i$.

2.2.3. Disease severity and individual benefits

Individual treatment benefits can be predicted/estimated using the estimated model parameters, not only for the subjects in the analysis but also for simulated new patients. The severity of a patient's outcome at a given time point is defined as the probability that the patient's outcome is outside of the therapeutic target (Diaz, 2016). The disease severity for patient i before treatment (time 0) is

$$s_{0,i} = 1 - \Phi\left(\frac{c - \boldsymbol{x}_{i1}^T\boldsymbol{\beta} - \boldsymbol{z}_{i1}^T\boldsymbol{\alpha}_i}{\sigma_e}\right)$$

where the therapeutic target is to achieve $y \leq c$.

The post-treatment severity of the patient at time t is

$$s_{t,i} = 1 - \Phi\left(\frac{c - \boldsymbol{x}_i^{(t)^T}\boldsymbol{\beta} - \boldsymbol{z}_i^{(t)^T}\boldsymbol{\alpha}_i}{\sigma_e}\right)$$

where $\boldsymbol{x}_i^{(t)}$ and $\boldsymbol{z}_i^{(t)}$ are covariate values measured at time t.

The individual benefit of the treatment for patient i after t units of time is defined as the reduction in disease severity at time *t* from time 0 (Diaz, 2016, 2019), that is,

$$b\left(t; \boldsymbol{\beta}, \boldsymbol{\alpha}_i, \boldsymbol{x}_i^{(t)}, \boldsymbol{z}_i^{(t)}\right) = s_{0,i} - s_{t,i}. \qquad (2)$$

2.2.5. Empirical Bayesian prediction of benefits

As described in 2.2, the individual treatment benefit is defined as the decrease of the disease severity from baseline for the patient (Diaz 2019). If $\widehat{\boldsymbol{b}}_i$ is the EB predictor of the patient's $\boldsymbol{b_i}$, the EB predictor of the individual benefit at time $t \geq 0$ is

$$b\left(t;\ \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}_i, \boldsymbol{x}_i^{(t)}, \boldsymbol{z}_i^{(t)}\right) = \left\{\Phi\left(\frac{c - \boldsymbol{x}_i^{(t)^T}\widehat{\boldsymbol{\beta}} - \boldsymbol{z}_i^{(t)^T}\widehat{\boldsymbol{\alpha}}_i}{\hat{\sigma}_e}\right) - \Phi\left(\frac{c - \boldsymbol{x}_{i1}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{z}_{i1}^T\widehat{\boldsymbol{\alpha}}_i}{\hat{\sigma}_e}\right)\right\} \times 100, \qquad (3)$$

where $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of $\boldsymbol{\beta}$ and $\hat{\sigma}_e$ is the maximum or restricted maximum likelihood estimate of the standard deviation of the pain score model residuals.

2.2.4. Transformation of outcome variable.

Here we are concerned with outcomes that decrease over time and have a minimum value during the study. If the outcome is continuous, we can use a logit transformation. If not, we can use a discrete logit transformation. For instance, if the outcome is in the range of 0 to m where m is the maximum value of the outcome. The following discrete logit transformation can be used to transform the discrete outcomes:

$$T(y_{ij}) = \log\left(\frac{y_{ij} + 1}{m + 1 - y_{ij}}\right) \quad (4)$$

where y_{ij} is the outcome for subject i at time j.

2.2.5. Application

In this study, we used EHR data from the Cerner HealthFacts® dataset (Cerner HealthFacts®; Kansas City, MO). The Cerner HealthFacts dataset is a deidentified EHR database, and this study exempted from institutional review by Western IRB (Olympia, WA). Adult patients undergoing spine

fusion surgery as inpatients in the United States between January 1, 2014 and December 31, 2015 were selected using International Classification of Diseases ICD-9 codes 81.00 to 81.08 and corresponding ICD-10 codes for spine fusion. Additional inclusion criteria were 1) patients with at least one pain score available on the day of surgery (day 0) and at least one pain score post-surgery; 2) the maximum baseline pain score was at least 7; and 3) patients had 1 to 5 days of post-surgical hospital stay. Patients without at least 6 months of records in the database prior to the surgery were excluded. We identified 940 patients who satisfied the inclusion criteria and the 330 patients from the hospital with the largest number of patients were selected as the subjects for this study to obtain greater homogeneity since each hospital may have different pain management protocols.

In the application, the outcome is maximum daily pain score post spine fusion surgery. The pain scores are patient-reported measurements that ranged from 0 to 10, with 0 indicating no pain and 10 indicating the most severe pain. The outcome of interest is the patient's maximum daily pain score, obtained at day 0 and during the 1-5 days of post-surgical hospital stay. In most cases, patients last pain score was observed on the day of the discharge. In a few cases, patients' pain score measurement was deaminated before the day of discharge. In these few cases, the outcome was considered censored on the day of the last pain score measurement. An offset of 0.01 was added to the LOS and censoring time to make them slightly larger than the time of the last pain scores. We used the jmre1 (Pantazis et al. 2010) command in Stata (StataCorp LLC, College Station, TX) for this analysis.

In the pain score model, the design matrix for the fixed effects is

$$\boldsymbol{X}_i^* = [\mathbf{1} \quad X_{i1} \quad X_{i2} \quad \boldsymbol{t}_i \quad X_{i3}]$$

where x_{i1}, x_{i2} denote the dichotomous variables Elderly (1 if age>65, 0 otherwise) and Depression (1 if the patient had a record of preoperative depression diagnosis, 0 otherwise), respectively. $X_{i1} = (x_{i1}, \dots, x_{i1})^T$, $X_{i2} = (x_{i2}, \dots, x_{i2})^T$, $\boldsymbol{t}_i = \left(t_{i1}, \dots, t_{in_i}\right)^T$ is the vector containing the days from surgery on which the pain scores were observed for subject i, and $X_{i3} = \left(x_{i2}t_{i1}, \dots, x_{i2}t_{in_i}\right)^T$ is the interaction between Depression and time.

In the LOS model, the design vector $\boldsymbol{x}_i^d = [1 \quad x_{i1} \quad x_{i2}]$, where x_{i1}, x_{i2} are Elderly and Depression variables, respectively.

The maximum daily pain score is 10, that is m=10 in formula (3). The discrete logit transformation we used for pain scores is

$$\log\left(\frac{y_{ij}+1}{11-y_{ij}}\right)$$

where y_i is the maximum daily pain score for subject i at Day_{ij}. The distribution of the original pain scores are highly skewed with higher frequencies for severe pain scores. After this transformation, the distributions of the EB predictor of the LOS model residuals and the random intercept and the random effect of time for the pain score model were relatively normal, suggesting good model fit.

The postoperative treatment target was defined as a maximum pain score level of 6 or lower, which corresponds to a transformed pain score of $c = T(6) = 0.3365$ or lower.

$\boldsymbol{\alpha}_{EB,i}^j$ was calculated using the Stata "predict" command after running the jmre1 command.

2.2.6. Monte Carlo computation of individual benefits based on patients' characteristics

As an alternative approach to analyze the pain management benefits, Monte Carlo computation was used to estimate the quartiles of the probability distribution of individual benefits for the four subpopulations by time (**Table 3**) using the algorithm below.

1. Draw 1000 random effects for each group of patients (depending on age and depression) from the distribution of the random effects
2. Generate random coefficients of the intercept and the time for the 1000 patients in each group by adding up the fixed effects and the random effects for the intercept and time, respectively
3. Calculate benefits for each patient on days 1 through 5 post-surgery using the benefit function by plugging in the therapeutic target of the transformed pain score, age, depression, the random coefficients, days post-surgery, and the variance of the pain score model residuals
4. Calculate median, p25, p75 for days 1 through 5 post-surgery for the 4 groups

2.2.7. Simulations to evaluate the performance of individual benefit predictions in hypothetical new patients

To evaluate how well the proposed method of benefit prediction would work in new patients, we assessed the predicted benefits through simulations for various prediction origins (t), prediction horizons (h) and distance of parameter estimates from true parameters (δ) in standard error units (Diaz 2019) (**Tables 4** and **5**).

The performance of the predictions was assessed using the approach used by Diaz (2019). Pearson's correlation (C_{t+h}) was used to examine the correlation between the predicted benefits and the simulated true benefits. Each value of C_{t+h} was computed with a simulated sample of 1000 new patients. The accuracy of the prediction was also assessed using relative bias (Diaz, 2019). Briefly, the bias was defined as the predicted benefit minus the true benefit for each simulated patient at each time point, and the relative bias ($\mathcal{B}_{t+h}$) was defined as {(mean of bias) / (mean of true benefit)} x 100 for each set of 1,000 simulated patients.

Simulations of the true and predicted benefits were carried out with the algorithm described below.

1. Define values of the distance δ between the parameter estimates in Table 1 and the true parameters, prediction origin t, prediction horizon h, age, and depression. Since t=0 represents baseline, and we also need a row for the drop-out model, a total of t+2 simulated responses are needed.
2. Obtain the estimates of both the vector of fixed effects and the variance of the random intercept from Table 1.
3. Calculate the true fixed effects for the pain score model and the true variance/covariance matrix of the random effects based on δ. For instance, $\delta = 0.5$ indicates each true parameter value is $\pm$ 0.5 times standard errors (SE) away from the corresponding parameter estimate of the model (Column 2 of Table 1). Hence, for 12 parameters, there are $2^{12} = 4096$ sets of true parameter values for a particular $\delta > 0$ (Diaz, 2019).

4. For SE of principle minor and determinant of D, multiply the estimates of principle minor and determinant by 0.6 and use as proxies for the SE.
5. Obtain the design matrix of dimension $(t+2) \times 8$ corresponding to the fixed effects. Also obtain the design matrix of dimension $(t+2) \times 3$ corresponding to the random effects.
6. Draw 1,000 random effects from a joint normal distribution with mean 0 and variance/covariance matrix D equal to the true variance calculated at step 4. These represent 1,000 new patients. Draw t+1 of error terms for each of the 1,000 patients.
7. Calculate the true benefits using the benefit formula (3), the true fixed effects from step #3, the simulated true random effect, and the errors from step 6.
8. Obtain EB predictors for the random effects using matrix calculation for the 1,000 new patients.
9. Calculate the predicted benefits using the benefit formula, the fixed effects from Table 1 and predicted random effects for the pain score model from step 8.

2.3. Results

2.3.1. The association between pain scores and LOS and the impact of depression and age on individual postoperative pain management

The joint model (**Table 3**) showed a positive covariance between the random residual of the LOS model and the random intercept of the pain score model (0.4985), between the random residual of the LOS model and the random effect of time in the pain score model (0.6631), as well as between the random intercept and the random effect of time in the pain score model (0.8471). Log-likelihood ratio test comparing the full model and the constrained model (setting the covariance between the random residual of the LOS model and the random intercept or the random effect of time of the pain score model) indicated that the association between the pain score model and the LOS model was significant (p = 1.697e-22).

As shown in Table 3, the preoperative depression comorbidity was significantly associated with higher pain scores at baseline on average (parameter estimate = 0.2278) whereas older age was significantly associated with lower baseline pain scores (parameter estimate = -0.1853). The interaction between

depression and time is positive (parameter estimate=0.1327), indicating that the slope of the decrease in pain scores overtime is less steep for patients with depression.

For the LOS model, older age was significantly associated with longer post-surgical LOS (parameter estimate = 0.2196), and preoperative depression comorbidity tended to have slightly longer LOS (parameter estimate = 0.0886).

2.3.2. Comparison of Empirical Bayesian quartile estimates with Monte Carlo estimates of individual benefits

We first analyzed the EB predicted individual pain management benefits in the 330 subjects during the study using formula (4). The medians and the first and third quartiles of the predicted benefits are shown in Table 2. During the 5 days' recovery post the surgery, all four groups of patients gradually received more benefits in pain management over time. For the same time points, how much benefits each subject received differ depending on the patient's preoperative depression status and age. The degree of variations in the amount of benefits within the same subgroup defined by depression and age also differ depending on the group and time points. For instance, in elderly patients with no depression, the median decrease in disease severity was 25.5% probability units compared to 12.1% for patients in the same age group with depression at day 1. The minimum and maximum benefits for elderly patients with no depression at day 1 were 4.4% and 34.2%, respectively, whereas for patients in the same age group with depression these were 3.5% and 19.1%, respectively. Similar effects of depression were observed in younger patients. The effects of age were more apparent in patients with depression, with younger patients showing lower median as well as minimum benefits at earlier time points.

As an alternative approach to analyze the pain management benefits, Monte Carlo computation was used to estimate the quartiles of the probability distribution of individual benefits for the four subpopulations (**Table 3**). The random effects (random residuals of the LOS model, random intercept and random slope of the pain score model) were simulated from a joint normal distribution with mean **0** and variance/covariance matrix (**Table 1**). The treatment benefits were calculated for the four subpopulations

at days 1 through 5 with formula (4), using the estimated values of the fixed effects shown in Table 1 in place of $\boldsymbol{\beta}$.

To visualize how the medians of the predicted pain management benefits changed over time post-surgery for the four groups of subjects in the study and compare the patterns with the Monte-Carlo computed benefits, we plotted the medians in Tables 3 and 4 in Panels A and B of **Figure 1**, respectively. Separate plots (**Figure 2**) were also made for each of the four groups, comparing EB-predicted and Monte-Carlo calculated benefits. The medians of benefits for patients increase at a slower pace for patients with depression compared to patients without depression in the same age group. In patients with no depression, the effects of age on the medians of benefits are minimal. In patients with depression, however, younger age was associated with slightly lower medians of benefits in earlier days post-surgery. The patterns for EB-predicted benefits are consistent with the Monte-Carlo computed benefits (**Figures 1 and 2**), suggesting that the medians of EB predictions, which are less computationally demanding than medians based on simulations, are good estimators of median benefits.

2.3.3. Evaluation of benefit prediction in simulated new patients using correlations between predictions and true benefits

Correlations between the predicted individual benefits and the true individual benefits in simulated new patients were analyzed the using Pearson's correlations (C_{t+h}). Each C_{t+h} was calculated from 1000 simulated patients in each of the four subpopulations defined by age and depression categories. Results for younger patients without depression are shown in **Table 4**. Results for the other three groups are included in the Supporting Information. Minimal correlations were observed when only baseline data were used for predictions (prediction origin $t = 0$). Correlations increased as t increased. This is true for predicting benefits for a given day (i.e., day 2 post-surgery) as well as predicting benefits for the same day (h=0), which is most apparent if the parameter estimates are the same as the true model parameters ($\delta = 0$). When parameter estimates moved further away from the true parameter values in the model, that is, as δ increased, the range of correlation values grew wider, as expected. However, the median of the correlations stayed approximately the same as for $\delta = 0$. When δ was sufficiently small and t was sufficiently large, the

correlations were good for the predictions not only when predicting the benefits achieved up to the current week ($h = 0$) but also for predictions of past benefits ($h < 0$). For predicting future benefits, correlations decreased as h increased, which is especially true for small t (i.e., t <2).

2.3.4. Evaluation of benefit prediction in simulated new patients using relative biases ($\mathcal{B}_{t+h}$)

To further evaluate the performance of the benefit predictor, we assessed the biases of the predicted benefits relative to the true benefits as defined in the Methods section. Each $\mathcal{B}_{t+h}$ was calculated from 1000 simulated patients for each of the four subpopulations, and the results for younger patients without depression are shown in **Table 5.** The negative signs indicate that the predicted benefits are smaller than the true benefits. Higher relative biases were observed when t=0. $\mathcal{B}_{t+h}$ decreased as t increased. This is true for predicting benefits for a given day (i.e., day 2 post-surgery) as well as predicting benefits for the same day (h=0), indicating that the more data we can use the less biased the prediction will be. When δ is sufficiently small, relative biases were relatively small when t > 0, suggesting relatively accurate predictions of past, current and future benefits when the patient provides at least one measure of the pain scores post-surgery. As expected, the range of possible values of $\mathcal{B}_{t+h}$ became wider as the δ increased. However, the median of $\mathcal{B}_{t+h}$ stayed approximately the same as for δ = 0.

Table 1. Mixed effects model of transformed pain scores from 330 subjects after spine fusion surgery.

Parameter name	Parameter estimate (SE)	p-value
Fixed effects for LOS (days)		
LOS intercept	0.2465 (0.0385)	<0.0001
Older age[a]	0.2196 (0.0644)	0.001
Depression[b]	0.0886 (0.0532)	0.096
Fixed effects for transformed pain score		
Pain score intercept	1.4704 (0.0544)	<0.0001
Older age[a]	-0.1853 (0.0931)	0.047

Depression[b]	0.2278 (0.0742)	0.001
Time (days)[c]	-0.6771 (0.0461)	<0.0001
Interaction between depression and time	0.1327 (0.0662)	0.045
Variance of random effects		
LOS residual, d_{11}	0.2281	--
Pain score intercept, d_{22}	0.1384	--
Time[b], d_{33}	0.0916	--
Covariances		
Cov (LOS residual, Pain score intercept), d_{12}	0.4985	--
Cov (LOS residual, Time), d_{13}	0.6631	--
Cov (Pain score intercept, Time), d_{23}	0.8471	--
Residual variance, σ^2	0.3835	--

SE: standard error.

[a]The dichotomous covariate was defined as 1 if the age of the subject was greater than 65, and 0 otherwise.

[b]The dichotomous covariate depression was defined as 1 if the subject had a record of depression diagnosis, and 0 otherwise.

[c]Time was defined as days post spine fusion surgery.

Table 2. Sample medians (and first and third quartiles) of individual benefits (x100) of postoperative pain management on days 1 through 5 for 330 subjects after spine fusion. Empirical Bayesian predictors of the subject's random effects were used for predicting treatment benefits, combining data with parameter estimates in Table 1.

Study group	Day 1	Day 2	Day 3	Day 4	Day 5
Age ≤ 65, no depression (N=191)	18.6 (9.3, 37.1)	59.3 (35.9, 81.1)	87.7 (70.9, 91.8)	93.0 (87.3, 95.5)	94.4 (89.7, 97.0)
Age ≤ 65, depression (N=80)	5.6 (1.3, 17.3)	23.8 (5.7, 56.1)	53.6 (16.2, 87.3)	81.3 (34.4, 95.2)	93.6 (57.2, 96.9)
Age > 65, no depression (N=49)	25.5 (4.4, 34.2)	64.2 (15.4, 76.3)	87.6 (34.6, 89.4)	89.4 (58.3, 93.3)	90.4 (76.2, 93.8)
Age > 65, depression (N=10)	12.1 (3.5, 19.1)	40.3 (12.2, 56.4)	72.5 (28.9, 83.5)	89.3 (52.6, 92.8)	91.6 (75.1, 96.1)

Table 3. Estimates of medians (and first and third quartiles) of individual benefits (x100) of postoperative pain management on days 1 through 5 after spine fusion, obtained with Monte Carlo computation. The model in Table 1 was used for simulating 1,000 patients for each study group.

Study group	Day 1	Day 2	Day 3	Day 4	Day 5
Age ≤ 65, no depression	18.5 (5.4, 39.8)	57.6 (21.6, 79.1)	82.8 (49.2, 90.6)	89.5 (72.1, 94.4)	92.4 (82.0, 96.3)
Age ≤ 65, depression	5.8 (1.3, 21.8)	28.5 (5.4, 65.4)	62.3 (14.7, 88.7)	85.1 (29.8, 94.5)	92.2 (51.1, 96.7)
Age > 65, no depression	25.1 (8.9, 46.2)	63.7 (30.3, 77.6)	80.1 (56.3, 87.8)	86.0 (71.7, 92.1)	88.9 (78.2, 93.8)
Age > 65, depression	11.0 (2.5, 28.9)	38.0 (9.2, 70.8)	69.4 (22.2, 87.4)	84.9 (40.5, 92.8)	89.7 (60.9, 94.9)

Table 4. Pearson correlations (C_{t+h}) between empirical Bayesian predicted benefits and true benefits of postoperative pain management on days 1, 3, 5 in simulated new patients who are under 65 and with no depression, by prediction origin (t), prediction horizon (h) and distance of parameter estimates from true parameters in standard error units (δ).

		$t+h$ (weeks)				
t $(weeks)$	δ	1	2	3	4	5
0	0	0.51	0.38	0.31	0.22	0.02
	0.2	0.45 (0.32, 0.56)	0.41 (0.25, 0.54)	0.34 (0.15, 0.50)	0.21 (0.04, 0.37)	0.00 (-0.02, 0.27)
	0.4	0.42 (0.16, 0.61)	0.36 (0.09, 0.56)	0.28 (-0.01, 0.52)	0.15 (-0.07, 0.40)	0.01 (-0.24, 0.38)
	0.8	0.31 (-0.28, 0.64)	0.22 (-0.37, 0.60)	0.12 (-0.32, 0.55)	0.05 (-0.14, 0.41)	0.06 (-0.28, 0.55)
	1.2	0.17 (-0.45, 0.65)	0.09 (-0.55, 0.62)	-0.02 (-0.40, 0.57)	0.03 (-0.14, 0.47)	0.08 (-0.30, 0.57)
1	0	0.69	0.69	0.63	0.59	0.42
	0.2	0.73 (0.62, 0.82)	0.69 (0.54, 0.79)	0.63 (0.45, 0.77)	0.54 (0.33, 0.68)	0.43 (0.19, 0.60)
	0.4	0.73 (0.49, 0.84)	0.68 (0.42, 0.82)	0.58 (0.29, 0.77)	0.49 (0.13, 0.71)	0.37 (0.01, 0.64)
	0.8	0.76 (0.02, 0.90)	0.70 (-0.17, 0.87)	0.61 (-0.24, 0.80)	0.47 (-0.18, 0.73)	0.35 (-0.10, 0.65)
	1.2	0.80 (-0.24, 0.93)	0.75 (-0.39, 0.89)	0.66 (-0.35, 0.82)	0.54 (-0.18, 0.76)	0.43 (-0.15, 0.70)
2	0	0.87	0.84	0.80	0.76	0.70
	0.2	0.87 (0.78, 0.93)	0.84 (0.78, 0.91)	0.80 (0.67, 0.88)	0.75 (0.57, 0.85)	0.69 (0.47, 0.81)
	0.4	0.88 (0.69, 0.94)	0.84 (0.63, 092)	0.79 (0.53, 0.89)	0.71 (0.31, 0.86)	0.64 (0.20, 0.83)
	0.8	0.91 (0.28, 0.96)	0.88 (0.04, 0.94)	0.81 (-0.14, 0.91)	0.72 (-0.11, 0.88)	0.64 (-0.03, 0.86)
	1.2	0.88 (0.06, 0.96)	0.84 (-0.10, 0.94)	0.77 (-0.14, 0.92)	0.69 (-0.11, 0.90)	0.62 (-0.09, 0.88)
3	0	0.93	0.92	0.89	0.88	0.83
	0.2	0.93 (0.87, 0.96)	0.91 (0.86, 0.95)	0.89 (0.80, 0.94)	0.86 (0.73, 0.93)	0.82 (0.66, 0.90)
	0.4	0.93 (0.79, 0.97)	0.92 (0.79, 0.96)	0.89 (0.67, 0.95)	0.85 (0.50, 0.93)	0.80 (0.35, 0.92)
	0.8	0.90 (0.30, 0.97)	0.90 (0.30, 0.96)	0.86 (0.05, 0.95)	0.81 (-0.03, 0.94)	0.76 (0.02, 0.93)
	1.2	0.89 (0.17, 0.98)	0.89 (0.33, 0.97)	0.84 (0.11, 0.96)	0.77 (0.01, 0.95)	0.71 (-0.01, 0.93)
4	0	0.95	0.95	0.94	0.92	0.90
	0.2	0.95 (0.91, 0.97)	0.95 (0.90, 0.97)	0.93 (0.87, 0.96)	0.91 (0.83, 0.96)	0.89 (0.76, 0.95)
	0.4	0.94 (0.84, 0.98)	0.95 (0.87, 0.97)	0.93 (0.77, 0.97)	0.90 (0.60, 0.96)	0.88 (0.46, 0.95)
	0.8	0.89 (0.28, 0.98)	0.92 (0.49, 0.98)	0.90 (0.20, 0.97)	0.86 (0.15, 0.97)	0.82 (0.00, 0.96)
	1.2	0.86 (0.06, 0.99)	0.90 (0.54, 0.98)	0.88 (0.27, 0.97)	0.84 (0.19, 0.97)	0.80 (-0.04, 0.96)
5	0	0.96	0.96	0.96	0.95	0.93
	0.2	0.96 (0.93, 0.98)	0.97 (0.93, 0.98)	0.96 (0.91, 0.98)	0.94 (0.86, 0.97)	0.93 (0.81, 0.97)
	0.4	0.94 (0.87, 0.98)	0.96 (0.89, 0.98)	0.95 (0.82, 0.98)	0.93 (066, 0.98)	0.91 (0.56, 0.97)
	0.8	0.88 (0.19, 0.99)	0.94 (0.62, 0.99)	0.93 (0.44, 0.98)	0.90 (0.25, 0.98)	0.86 (-0.02, 0.98)
	1.2	0.83 (-0.02, 0.99)	0.90 (-0.57, 0.99)	0.89 (0.40, 0.98)	0.87 (0.31, 0.98)	0.84 (0.106,0.97)

Note: Parameter estimates in the second column of Table 1 were used for predicting treatment benefits. δ = 0 corresponds to the ideal situation when parameter estimates are equal to the true model parameters, in which case there is only one C_{t+h}. For $\delta > 0$, each entry in the table gives the median (minimum, maximum) of 4096 (=2^{12}) values of C_{t+h} corresponding to different combinations of parameter values that are at a distance of δ standard errors from their corresponding estimates. Each value of C_{t+h} was computed using 1,000 simulated new patients who were cannabis users.

Table 5. Relative biases ($\mathcal{B}_{t+h}$) of empirical Bayesian predictions of postoperative pain management benefits on days 1, 3, 5 in simulated new patients who are under 65 and with no depression, by prediction origin (t), prediction horizon (h) and distance of parameter estimates from true parameters in standard error units (δ).

		$t+h$ (weeks)				
t $(weeks)$	δ	1	2	3	4	5
0	0	-13.5	13.8	25.0	24.6	20.5
	0.2	-15.6 (-28.7, -1.7)	12.9 (3.9, 24.3)	25.0 (12.0, 39.5)	22.6 (10.6, 38.3)	18.6 (6.9, 33.5)
	0.4	-16.1 (-35.5, 7.9)	14.2 (-1.9, 28.8)	30.2 (3.5, 50.3)	29.1 (3.7, 52.5)	24.9 (3.4, 50.0)
	0.8	-20.6 (-47.8,41.3)	18.2 (-10.8, 40.9)	42.4 (-5.2, 74.4)	44.9 (-0.7, 82.4)	41.0 (-0.1, 78.0)
	1.2	-26.7 (-55.2,97.9)	22.3 (-14.1, 60.8)	53.4 (-4.6, 99.4)	58.0 (1.4, 123.4)	55.2 (2.9,117.6)
1	0	-9.4	9.8	14.5	13.8	11.0
	0.2	-9.0 (-20.2, 2.6)	7.7 (0.0, 16.3)	15.3 (5.8, 26.3)	15.8 (5.4, 27.4)	13.5 (4.6, 25.9)
	0.4	-8.9 (-26.7, 11.8)	8.4 (-3.8, 11.0)	18.3 (-1.0, 33.4)	20.1 (0.0, 36.3)	19.0 (0.5, 38.8)
	0.8	-12.3 (-35.0,37.1)	10.0 (-10.4, 29.4)	24.8 (-8.2, 49.5)	22.9 (-3.4, 61.4)	30.6 (-0.5, 57.5)
	1.2	-17.5 (-40.1,76.7)	10.9 (-10.7, 40.0)	28.2 (-7.4, 59.4)	35.8 (-1.8, 17.9)	38.8 (1.5, 71.5)
2	0	-4.5	4.3	9.9	9.0	8.5
	0.2	-4.5 (-12.4, 3.5)	4.0 (-1.5, 9.9)	8.1 (1.8, 13.1)	9.0 (2.7, 16.6)	8.5 (2.2, 16.3)
	0.4	-3.8 (-16.8, 8.8)	3.7 (-4.2, 13.3)	9.0 (-2.9, 16.0)	11.2 (-1.0, 19.4)	11.6 (-0.2, 21.5)
	0.8	-5.9 (-28.3, 29.5)	2.9 (-14.0, 21.2)	9.8 (-6.5, 29.7)	13.3 (-1.7, 35.0)	14.8 (-0.2, 35.9)
	1.2	-10.5 (-47.4,62.7)	1.2 (-34.9, 27.8)	9.8 (-27.6, 36.7)	13.7 (-22.5, 42.5)	15.3 (-21.0, 47.0)
3	0	-3.6	1.4	5.0	4.0	4.5
	0.2	-2.1 (-8.5, 5.4)	2.0 (-2.7, 8.3)	4.1 (0.9, 7.6)	4.9 (1.0, 8.8)	5.0 (1.0, 9.5)
	0.4	-1.0 (-13.8, 8.9)	1.4 (-7.3, 9.6)	3.8 (-2.3,11.2)	5.1 (-0.9, 11.9)	5.9 (-0.4, 12.5)
	0.8	-3.6 (-31.4, 27.4)	-0.9 (-23.7, 17.0)	3.3 (-17.7, 18.8)	5.6 (-12.8, 23.2)	6.6 (-11.0, 25.3)
	1.2	-8.3 (-57.2, 64.4)	-4.7 (-47.5, 27.5)	1.1 (-42.8, 23.8)	4.5 (-39.2, 27.4)	6.3 (-37.0, 34.1)
4	0	0.7	1.3	2.1	3.4	2.9
	0.2	-0.8 (-6.6, 5.6)	1.0 (-4.5, 5.7)	2.2 (-1.4, 5.8)	2.6 (0.0, 6.0)	2.8 (0.6, 5.7)
	0.4	0.0 (-13.4, 10.3)	0.2 (-9.8, 9.6)	1.6 (-6.4, 8.0)	2.3 (-4.0, 9.2)	2.7 (-2.8, 10.0)
	0.8	-3.1 (-34.0, 28.0)	-3.2 (-27.4, 15.9)	-0.3 (-23.8, 12.6)	1.9 (-21.3, 15.3)	3.0 (-18.7, 16.6)
	1.2	-7.9 (-61.1, 66.3)	-8.8 (-54.6, 23.7)	-4.8 (-49.1, 19.1)	-1.9 (-46.2, 18.6)	-0.1 (-45.2, 22.3)
5	0	0.0	0.7	1.2	1.5	2.1
	0.2	-0.3 (-0.9, 5.7)	0.5 (-5.3, 5.5)	1.2 (-2.9, 4.8)	1.6 (-1.7, 4.9)	1.7 (-1.2, 4.4)
	0.4	0.1 (-12.7, 11.1)	-0.4 (-12.1, 8.8)	0.4 (-9.1, 7.6)	1.1 (-6.9, 6.8)	1.5 (-5.4, 7.8)
	0.8	-2.6 (-35.9, 31.1)	-4.6 (-31.5, 14.8)	-2.4 (-26.9, 11.5)	-0.4 (-25.1, 10.6)	0.7 (-22.3, 12.1)
	1.2	-7.9 (-63.6, 72.1)	-11.3 (-57.3,23.2)	-8.1 (-52.8, 17.6)	-5.9 (-51.3, 15.5)	-4.2 (-48.8, 17.5)

Note: Parameter estimates in the second column of Table 1 were used for predicting treatment benefits. $\delta = 0$ corresponds to the ideal situation when parameter estimates are equal to the true model parameters, in which case there is only one $\mathcal{B}_{t+h}$. For $\delta > 0$, each entry in the table gives the median (minimum, maximum) of 4096 ($=2^{12}$) values of $\mathcal{B}_{t+h}$ corresponding to different combinations of parameter values that are at a distance of δ standard errors from their corresponding estimates. Each value of $\mathcal{B}_{t+h}$ was computed using 1,000 simulated new patients who were cannabis users.

Figure 1. Comparison of estimators of medians of individual pain management benefits at days 1 through 6. (A) Plots of medians of predicted antipsychotic treatment benefits for the 330 subjects in this study. (B) Medians of the individual benefits of 1,000 patients in each of the four groups were computed using Monte-Carlo simulations assuming the Model in Table 1.

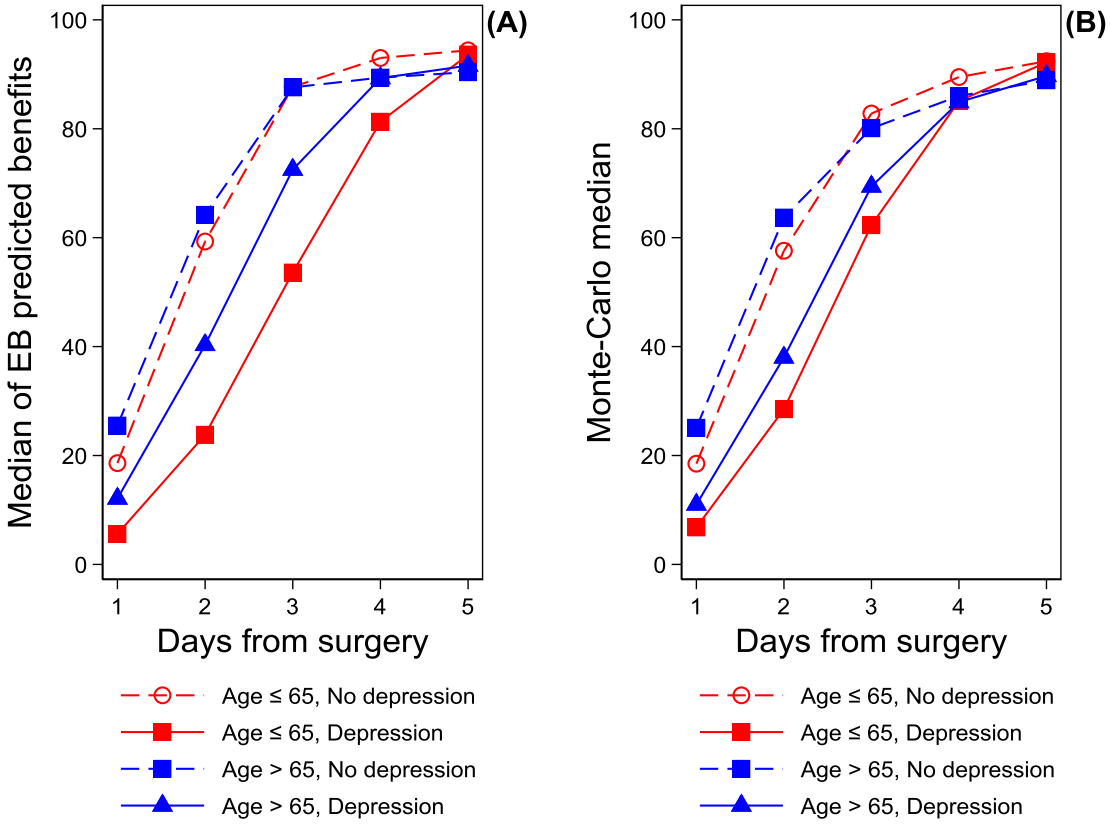
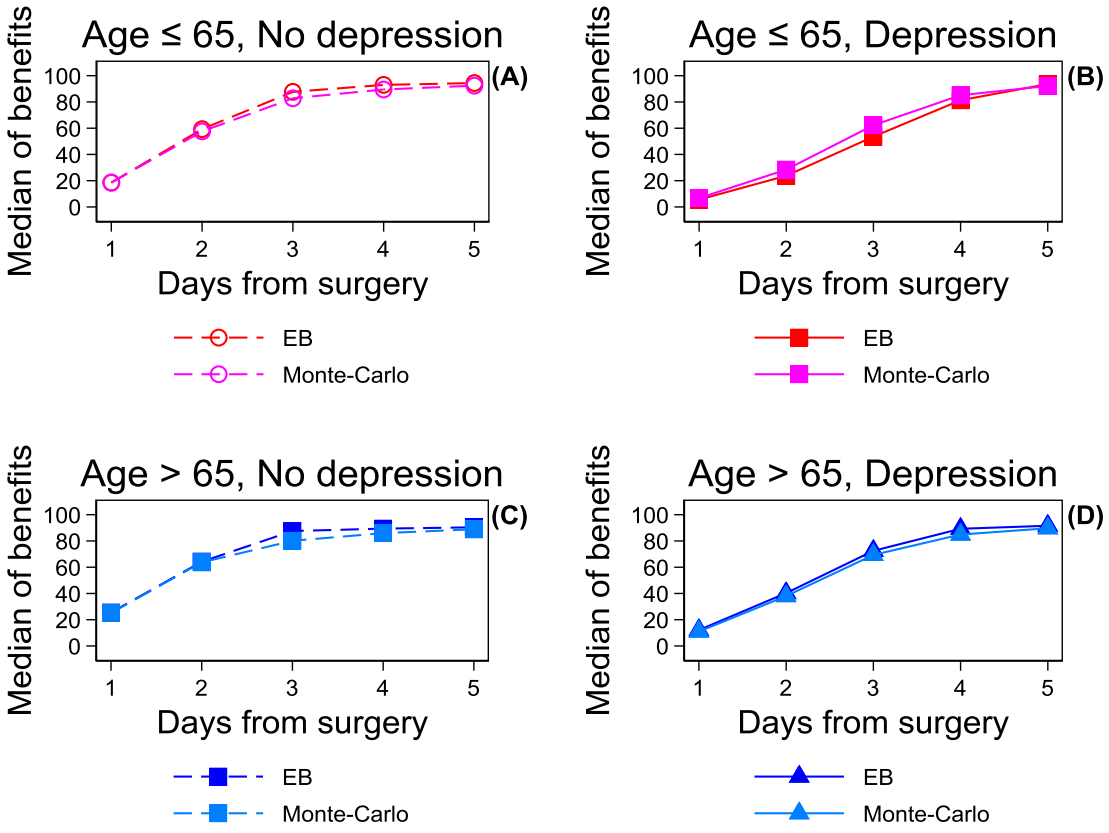


Figure 2. Comparisons of medians of empirical Bayes-predicted and Monte-Carlo computed antipsychotic treatment benefits for the four groups of patients. Medians of the empirical Bayes-predicted treatment benefits were calculated for the 330 subjects in this study, and medians of the Monte-Carlo computed benefits were calculated using 1,000 simulated patients for each of the four groups assuming the Model in Table 1. (A) Younger age with no depression. (B) Younger age with depression. (C) Older age with no depression. (D) Older age with depression.



2.4. Discussion

In this paper, we extended the methods for individual treatment benefit prediction using mixed-effects models proposed by Diaz (2016, 2019) to allow non-ignorable missingness in the longitudinal data. Although modeling informative drop-out in the analysis of clinical trial data with some patients dropping out of the study after randomization is not new, the idea of extending this concept to real-world hospital data for which the follow-up data are incomplete due to hospital discharge is novel. This is the first paper to analyze individual treatment benefits using EHR data. Since RWD are becoming more and more

important in clinical evidence generation, this offers a new way of analyzing treatment effects from personalized medicine perspective.

In the application, longitudinal pain score data of patients undergoing spine fusion surgery, extracted from Cerner HealthFacts® HER database, were modeled simultaneously with post-surgical LOS using a 2-PM model for joint mixed effects. This is another novelty of the current study. Previously, the 2-PM model was defined as a mixed effects model of a single outcome of repeated measures with random effects for both the intercept and time-dependent covariates. By extending the 2-PM model to allow for joint modeling of the outcome of interest and the hospital LOS, the prediction of individual treatment benefits can now handle longitudinal data with non-ignorable missingness. The correlation between the longitudinal outcome and the LOS is taking into consideration, leading to more reliable and accurate estimation of the model parameters. More accurate parameter estimates in turn lead to better prediction of individual treatment benefits.

Although the "predict" command in JMRE1 provides the EB estimates of the random effects in the joint mixed effects model, previous publications by the author did not explicitly describe the method used in the prediction (Touloumi et al. 1999, Pantazis et al. 2010). In this study, we provided detailed information on the calculation of the random effects using matrix algebra. The joint model defines that the level 1 residual in the LOS model is always 0; therefore, the values of LOS model level 1 residuals as well as the variance were set to 0.

We are interested in longitudinal outcomes that decrease over time and stabilize at a minimum value. A transformation of the outcome is necessary to make sure the estimated outcomes are within a meaningful range. For discrete outcomes, like the pain scores we used in the application, a discrete logit transformation can be used as described in the Methods section. This is another contribution of the current study. To our best knowledge, there has been no publications in the literature that proposed this type of transformation for discrete outcomes such as patient-reported scores.

A lognormal accelerated failure time model was used to model the post-surgical hospital LOS. This is a reasonable model in terms of the pattern of the hazard of discharge. We are investigating patients

undergoing a major surgery, who would be unlikely to go home immediately after surgery and more likely to go home in the several days after the surgery, but then less likely to go home if they remain in the hospital longer due to any complications associated with the surgery.

We used two methods to estimate the individualized pain management benefits in the application. The results from the EB method are quite similar to those from Monte-Carlo computations. This confirms that the prediction of individualized benefits using EB-predicted random effects is reliable. In JMRE1, the "predict" command gives the EB-predicted random effects. The makes it convenient for researchers to implement the prediction of individualized benefits for their data using the Stata.

The prediction performance using the 2-PM model for joint mixed effects was evaluated using Pearson's correlation and relative bias comparing predicted benefits with true benefits for simulated new patients. Results showed that, except when only baseline data are available, the prediction of benefits is reliable, with small median relative biases and good correlations when the model parameter estimates are reasonably close to the true parameter values. When the model parameter estimates move further away from the true parameter values, the range of the predicted correlation or relative biases get wider, especially when the prediction origin t is small. As the prediction origin t goes larger, the results become more stable and less sensitive to the changes in δ.

In summary, we proposed to use a 2-PM model with joint mixed effects that simultaneously models the longitudinal outcome and the hospital LOS for predicting individualized treatment benefits using unbalanced continuous or discrete outcomes in EHR data. Evaluation of the prediction using simulations demonstrated that the prediction is reliable in the application used in this study, given that the parameter estimates are not far from the true parameter values. This method can be used to analyze individualized benefits for many longitudinal clinical outcomes in the EHR data. The JMRE1 command is conveniently available in Stata, making it practical for the application of this method.

2.5. References

de Leon, J. (2012). Evidence-based medicine versus personalized medicine: are they enemies? *Journal of Clinical Psychopharmacology* 32:153-164.

Ruberg, S.J., Chen, L., Wang, Y. (2010). The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. *Clinical Trials* 7:574-583.

Andrews, N., Cho, H. (2018). Validating effectiveness of subgroup identification for longitudinal data. *Statistics in Medicine* 37:98-106.

Botts, S., Diaz, F. J., Santoro, V., Spina, E., Muscatello, M. R., Cogollo, M., Castro, F. E., de Leon, J. (2008). Estimating the effects of co-medications on plasma olanzapine concentrations by using a mixed model. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 32:1453-1458.

Cho, H., Wang, P., Qu, A. (2017). Personalized treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica* 27:187-205.

Diaz, F.J. (2016). Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. *Statistics in Medicine* 35:4077-4092.

Diaz, F.J. (2019). Estimating individual benefits of medical or behavioral treatments in severely ill patients. *Statistical Methods in Medical Research* 28:911-927.

Diaz, F.J., Rivera,T.E., Josiassen, R.C., de Leon, J. (2007). Individualizing drug dosage by using a random intercept linear model. *Statistics in Medicine* 26:2052-2073.

Diaz, F.J., Santoro, V., Spina, E., Cogollo, M., Rivera, T.E., Botts, S., de Leon, J. (2008) Estimating the size of the effects of co-medications on plasma clozapine concentrations using a model that controls for clozapine doses and confounding variables. *Pharmacopsychiatry* 41:81-91.

Diaz, F.J., de Leon, J. (2013a) The mathematics of drug dose individualization should be built with random effects linear models. *Therapeutic Drug Monitoring* 35:276-277.

Diaz, F.J., Berg, M.J., Krebill, R., Welty, T., Gidal, B.E., Alloway, R., Privitera, M. (2013b). Random-effects linear modeling and sample size tables for two special crossover designs of average bioequivalence studies: the four-period, two-sequence, two-formulation and six-period, three-sequence, three-formulation designs. *Clinical Pharmacokinetics* 52:1033-1043.

Diaz, F.J., de Leon, J. (2013a) The mathematics of drug dose individualization should be built with random effects linear models. *Therapeutic Drug Monitoring* 35:276-277.

Diaz, F.J., Berg, M.J., Krebill, R., Welty, T., Gidal, B.E., Alloway, R., Privitera, M. (2013b). Random-effects linear modeling and sample size tables for two special crossover designs of average bioequivalence studies: the four-period, two-sequence, two-formulation and six-period, three-sequence, three-formulation designs. *Clinical Pharmacokinetics* 52:1033-1043.

Diaz, F.J., Eap, C.B., Ansermot, N., Crettol, S., Spina, E., de Leon, J. (2014) Can valproic acid be an inducer of clozapine metabolism? *Pharmacopsychiatry* 47:89-96.

Senn, S. (2016). Mastering variation: variance components and personalised medicine. *Statistics in Medicine* 35:966-977.

Zhu, X., Qu, A. (2016). Individualizing drug dosage with longitudinal data. *Statistics in Medicine* 5:4474-4488.

Gewandter, J. S., McDermott, M. P., He, H., Gao, S., Cai, X., Farrar, J. T., et al. (2019). Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine. *Clinical Pharmacology & Therapeutics* 106:204-210.

Miksad, R.A., Abernethy, A.P. (2018) Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. *Clinical Pharmacology & Therapeutics* 103:202-205.

Little R.J.A., Rubin D.B. (2002) *Statistical Analysis with Missing Data, Second Edition*. Wiley, New York.

Pantazis, N. and Touloumi, G. (2010). Analyzing longitudinal data n the presence of informative drop-out: The jmre1 command. *Stata Journal* 10: 226-251.

Hedeker, D., Gibbons, R.D. (2006). *Wiley Series in Probability and Statistics. Longitudinal Data Analysis.* Hoboken, NJ: Wiley-Interscience.

Laird N.M. (1998). Missing data in longitudinal studies. *Statistics in Medicine*. 7:305-315.

Touloumi, G., Pocock, S.J., Babiker, A.G., Darbyshire, J.H. (1999). Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. *Statistics in Medicine* 18:1215-1233.

Schluchter, M.D. (1992) Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* 11:1861-70.

De Gruttola, V., Tu, X.M. (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* 50:1003-1014.

Touloumi, G., Pocock, S.J., Babiker, A.G., Darbyshire, J.H. (1999). Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. *Statistics in Medicine* 18:1215-1233.

Crowther, M.J., Abrams, K.R., Lambert, P.C. (2012) Flexible parametric joint modelling of longitudinal and survival *Statistics in Medicine* 31:4456-4471.

Armero, C., Forte, A., Perpiñán, H., Sanahuja, M.J., Agustí, S. (2018) Bayesian joint modeling for assessing the progression of chronic kidney disease in children. *Statistical Methods in Medical Research*. 27:298-311.

Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R. (2018). joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodoly*. 18:50.

Shardell, M., Ferrucci, L. (2018) Joint mixed-effects models for causal inference with longitudinal data. *Statistics in Medicine* 37:829-846.

Schluchter, M.D., Piccorelli A.V. (2019). Shared parameter models for joint analysis of longitudinal and survival data with left truncation due to delayed entry - Applications to cystic fibrosis. *Statistical Methods in Medical Research*. 28:1489-1507.

Papageorgiou, G., Mauff, K., Tomer, A., Rizopoulos, D. (2019). An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes. *Annual Review of Statistics and Its Application*. 6:223-240

Gronski L., Martinson W., Singh K.P., Ryan J. (2012). Utility of daily troponin orders for identifying acute myocardial infarction patients for quality improvement. *Critical Pathway in Cardiology*. 11:74-76.

Edelstein C.L. (2008). Biomarkers of acute kidney injury. Advances in Chronic Kidney Disease. 15:222-234.

Shaw A.D., Mythen MG., Shook D., Hayashida D.K., Zhang X., Skaar J.R., Iyengar S.S., Munson S.H. (2018). Pulmonary artery catheter use in adult patients undergoing cardiacsurgery: a retrospective, cohort study. Perioperative Medicine. 7:24.

Chapter 3: Effects of depression and age on individual benefits of pain management post spinal fusion: an analysis of longitudinal hospital data

In collaboration with Drs. Nikos Pantazis, Ph.D., Jose de Leon, M.D., and Francisco J. Diaz, Ph.D.

(See collaborator affiliations in Acknowledgements)

3.0. Abstract

Objectives: This study analyzed the impact of depression and age on individual benefits of postoperative pain management in lumbar spinal fusion patients using longitudinal observational data.

Methods: Cerner HealthFacts electronic health records were used. Patients were selected using International Classification of Diseases (ICD)-9 codes and ICD-10 codes for spinal fusion and predetermined inclusion/exclusion criteria. A joint multivariate mixed model of pain scores and length of hospital stay was used to analyze individual benefits.

Results: Depression was significantly associated with higher baseline pain scores ($p=0.001$) on average, whereas geriatric age was associated with lower baseline pain scores ($p=0.047$). Antidepressant use had no significant effects on postoperative pain scores in patients with depression. Although pain management benefits tended to increase with time, the amount and rate of change of the benefits depended on depression status and age. More patients with depression received small benefits than those without depression after controlling for age and time. For patients with depression, non-geriatric age was associated with slower individual benefits development, except for those achieving the highest benefits. In general, the detrimental depression effects on individual benefits outweigh age effects in the patients achieving the highest benefits. Patients with higher immediate benefits tended to have shorter lengths of stay.

Conclusions: This study revealed that preoperative depression and geriatric age may be important factors affecting individual benefits of postoperative pain management in patients undergoing spinal fusion surgery. Depression had a negative impact on pain relief, while age had varied effects depending on depression status and other traits.

Keywords: individual benefits, random effects, depression, pain management, spinal fusion.

3.1. INTRODUCTION

Although lumbar spinal fusion is the top procedure for treating chronic low back pain and is the second most common low back operation overall, better understanding is needed of how patients' characteristics influence postoperative outcomes (Gaudin et al., 2017). Depression is known to be associated with chronic pain such as back pain (Trivedi, 2004) and is a negative predictor of spinal fusion outcomes (Gaudin et al., 2017). Retrospective cohort studies have found that: 1) patients with pre-existing depression were absent from work for more days after spinal fusion surgery compared to those without depression (Anderson et al., 2015), and 2) preoperative depression influences patient satisfaction independent of the surgery's effectiveness (Adogwa et al., 2013).

Patient-reported maximum pain levels on a scale from 0 to 10 are often used as postoperative quality measures to monitor pain relief and track patients' progress after spinal fusion. Studies of risk factors for severe postoperative pain have provided varying results. The risk factors could be procedure-specific; however, preoperative chronic pain and younger age were associated with higher postoperative pain level independent of the type and extent of the surgery in pooled data from 150 German hospitals (Gerbershagen et al., 2014). In a German registry of knee replacement, older age was associated with lower reported maximum pain levels. On the other hand, the elderly patients did not report less functional impairment caused by pain, suggesting that they tend to underreport their pain levels (Weinmann et al., 2017).

It is important to further understand the impact of patients' characteristics such as preoperative depression and age on individual benefits of pain management after spinal fusion surgery. Generalized linear mixed-effects modeling is a statistical approach useful for predicting individuated treatment benefits (Diaz, 2016 and 2019), which take into consideration the heterogeneity of patients' characteristics including unknown traits. While traditional statistical analyses focus on average treatment effects, mixed-effects modeling can analyze the variation of treatment effects in individual patients.

Electronic health records (EHR) provide valuable resources for longitudinal studies and understanding risk factors associated with poor clinical outcomes. However, they may not provide complete follow-up, and the missing data are not at random since hospital discharge may depend in part on expected

but unrecorded clinical outcomes after discharge (Ibrahim and Molenberghs, 2009). This is called "non-ignorable missingness" and requires novel statistical techniques.[14] Ignoring the unbalanced nature of longitudinal EHR data may lead to serious bias (Albers et al., 2018).

In this study, we use novel statistical methods to evaluate the effects of depression and geriatric age (age>65 years) on patient-reported pain levels (Diaz, 2016 and 2019; Pantazis and Touloumi, 2010). The main goal is to measure and compare individual benefits of postoperative pain management, using EHR data from patients undergoing spinal fusion surgeries (Cerner HealthFacts®; Kansas City, MO).

3.2. METHODS

Data source and study subjects

The EHR dataset (Cerner HealthFacts®, Kansas City, MO) is deidentified and has been used in previously published articles (Shaw et al., 2018; Urman et al., 2018). An Institutional Review Board (IRB) exemption for this study was granted by Western IRB (Olympia, WA). We selected adult inpatients undergoing spinal fusion surgery in the United States between January 1, 2014, and December 31, 2015, using International Classification of Diseases ICD-9 codes 81.00 to 81.08 and corresponding ICD-10 codes.

Additional inclusion criteria were 1) at least one pain score on the day of surgery (day 0) and at least one score after that day; 2) a maximum score on day 0 between 7 and 10 inclusive; 3) 1 to 5 days post-surgical hospital stay; and 4) at least 6 months of history captured in the database prior to the surgery. For greater sample homogeneity, patients from the hospital with the largest number of patients meeting the above criteria were selected for this study. The reason for choosing a single hospital is that each hospital may have different pain management protocols. We identified 940 patients who satisfied the inclusion criteria, and 330 from the hospital with the largest number of patients were selected (Table 1).

Pain assessments

The numerical patient-reported pain scores ranged from 0 to 10, with 10 indicating the most severe pain (0: no pain, 1-3: mild pain, 4-6: moderate pain, 7-10 severe pain). The outcome of interest was the patient's maximum daily score, obtained at day 0 and during 1 to 5 days of post-surgical hospital stay (Table

2). Since patients' pain levels were not measured after discharge, this longitudinal observational study conveys the challenges of a highly unbalanced dataset caused by non-random missing data. Since pain scores are usually lower on or after the discharge day, the assumption that missing data would be random, which is required by standard longitudinal statistical models, is violated (Ibrahim and Molenberghs, 2009).

Depression assessments

Depression comorbidity was defined as having ICD-9 codes (3004, 30112, 3090, 3091, and 311) or ICD-10 codes (F320, F321, F322, F323, F328, F3281, F3289, F329, F330, F331, F332, F333, F338, F339, F341, F4321) during the hospital stay or within 6 months before admission, or having received antidepressants during the stay.

Hospital length of stay (LOS)

The patients' hospital LOS after surgery may be affected by their characteristics and responses to postoperative pain management. Pain levels are usually not measured after discharge and even when patients are in the hospital their pain measurements may be terminated for various reasons. Ignoring this incomplete follow-up in the data analysis could lead to serious bias (Ibrahim and Molenberghs, 2009). Hence, there is a need to apply special methods that account for the relationship between pain scores and LOS and model the premature termination of measurements in some patients.

Statistical model

This study utilized a joint multivariate random-effects (JMRE) model (Touloumi et al., 1999; Pantazis and Touloumi 2010), which is a generalized linear mixed-effects model that accounts for non-ignorable missingness. The model combined a model of daily maximum pain scores with a model of LOS (Table 3, Footnotes a-e). The daily maximum pain scores were transformed to improve the model's goodness-of-fit.

The variables included in the pain model were older age (1 if age >65 years, 0 otherwise), depression (1 if the patient had depression comorbidity, 0 otherwise), time as the number of days from surgery, and the interaction between depression and time. The transformed pain scores followed a linear

time trend. The intercept and the time slope were considered random, meaning they were different for each patient (Diaz, 2016 and 2019).

Details on the LOS model are provided in Table 3, Footnote c. It was assumed that the random residual of the LOS model was correlated with both the random intercept and random time slope of the pain model. Initial explorations showed that gender and race had no significant effects on either the pain scores or LOS and were therefore not included in the final model.

Individual pain management benefits

The severity of the patient's disease is defined as the probability of being outside the pain treatment target, which in turn is defined as a daily maximum pain score ≤6 (Table 4, Footnote a) (Diaz, 2016 and 2019). The patient's individual treatment benefit is defined as the decrease in disease severity from baseline (x100).

To examine how much benefit patients received from postoperative pain management during the 5 days after spinal fusion, we predicted the individual benefits for each of the 330 patients. Estimated random effects for each patient were used to predict treatment benefits, combining all available patient data with parameter estimates in Table 2. Details regarding calculation of the empirical Bayes (EB) predictors of the benefits are provided in Table 4, Footnote a (Diaz, 2016 and 2019).

For each patient, individual benefits were predicted from day 0.2 to day 5 by 0.2-day increments. Although patients' pain scores were observed for days 0 to 5, benefits can be predicted for any non-integer interval from 0 to 5 days using the formula in Table 4, Footnote a (Diaz, 2016 and 2019). Median, 25^{th} and 75^{th} percentiles of individual benefits were calculated. For each of the 4 groups determined by age and depression status, these statistics were plotted (Figure 1) and presented in Table 4 for days 1 through 5. To compare the evolution of individual benefits over time across the 4 groups, we plotted histograms of the benefits (Figure 2).

3.3. RESULTS

Patient Characteristics

Patients' characteristics, pain medications and antidepressant medications are described in Table 1. Almost half (46%) the patients had comorbid depression. Depression was more frequent in females (54%, 94/173) than in males (36%, 56/157) and in non-geriatric patients (49%, 132/271) than in geriatric patients (31%,18/59). Baseline pain scores and hospital LOS are in Table 2.

Table 1. Demographics and clinical characteristics of 330 patients who underwent a spinal fusion surgery.

	Mean	SD
Age (years)	53.9	12.4
	%	
GERIATRIC AGE (>65 years)		
Yes	18 (59/330)	
No	82 (271/330)	
GENDER		
Female	52 (173/330)	
Male	48 (157/330)	
RACE		
Caucasian	93 (308/330)	
African American	2 (7/330)	
Other	5 (15/330)	
PAIN MEDICATION		
Opioids and acetaminophen	78 (257/330)	
Opioids, NSAIDs and acetaminophen	18 (60/330)	
Opioids only	4 (12/330)	
Opioids and NSAIDs	<1 (1/330)	
DEPRESSION		
Yes	46 (150/330)	
No	54 (180/330)	
ANTIDEPRESSANT MEDICATION		
Taking antidepressants	81 (121/150)	
SSRI	34 (51/150)	
SNRI	11 (17/150)	
Other[a]	9 (13/150)	
SSRI and other	7 (11/150)	
SSRI and TCA	5 (8/150)	
TCA	5 (7/150)	
SNRI and TCA	4 (6/150)	
SSRI and SNRI	1 (2/150)	
SNRI and other	1 (2/150)	
MAOI	<1 (1/150)	
Other and TCA	<1 (1/150)	
SSRI, SNRI and other	<1 (1/150)	
SSRI, other and TCA	<1 (1/150)	

Abbreviations: SD = standard deviation, MAOI = monoamine oxidase inhibitor, NSAID = nonsteroidal anti-inflammatory drug, SNRI = serotonin norepinephrine reuptake inhibitor, SSRI = serotonin selective reuptake inhibitor, TCA = Tricyclic antidepressant.

[a]The EHR database did not itemize the medications in the "Other" category.

Joint model and the impact of depression and age on pain scores and LOS

We found positive correlations between 1) high baseline pain scores and longer postoperative LOS ($r=0.50$, $p<0.001$), 2) slower pain reduction and longer LOS ($r=0.67$, $p<0.001$), and 3) high baseline pain

scores and slower pain reduction post-surgery (r=0.85, p<0.001). These significant correlations indicated that 1) patients who had higher baseline pain scores tended to stay longer after surgery; 2) patients whose pain decreased more slowly after surgery tended to stay longer; and 3) patients with higher pain scores at baseline tended to have slower pain reduction after surgery.

The pain model demonstrated that, on average: 1) a preoperative record of depression was significantly associated with higher baseline pain scores (P = 0.001; Table 3); 2) geriatric age was significantly associated with lower baseline pain scores (P = 0.047); 3) a significant interaction existed between depression and time (parameter estimate=0.1327, p=0.045), meaning that patients with depression had significantly slower pain reduction after surgery.

The LOS model demonstrated that, on average: 1) geriatric age was significantly associated with longer LOS (p = 0.001; Table 3); and 2) depression tended to be associated with a slightly longer LOS, although it did not reach significance (P = 0.096).

Table 2. Mean and SD of stratified maximum baseline pain scores and hospital LOS after surgery in 330 patients who underwent a spinal fusion surgery

	Baseline Pain scores		LOS			
	Mean	SD	Mean	SD	Min	Max
All (N=330)	8.65	1.11	1.62	1.00	1.0	5.0
GERIATRIC AGE[a]						
Yes (N=59)	8.31	0.99	1.98	1.17	1.0	5.0
No (N=271)	8.72	1.12	1.54	0.94	1.0	5.0
GENDER						
Female (N=173)	8.66	1.11	1.64	1.03	1.0	5.0
Male (N=157)	8.62	1.11	1.59	0.97	1.0	5.0
RACE						
Caucasian (N=308)	8.62	1.11	1.61	1.00	1.0	5.0
African American (N=7)	8.71	1.25	1.71	1.50	1.0	5.0
Other (N=15)	9.20	1.21	1.67	0.62	1.0	5.0
DEPRESSION						
Yes (N=150)	8.81	1.13	1.67	1.03	1.0	5.0
No (N=180)	8.51	1.07	1.57	0.97	1.0	5.0

Abbreviations: SD = standard deviation, LOS = Length of stay.
[a]Geriatric age was defined as age >65 years.

Table 3. Joint random-effects model of transformed daily maximum pain scores and hospital length of stay from 330 patients after spinal fusion surgery

Parameter name	Estimate	P	95% CI
FIXED EFFECTS FOR TRANSFORMED PAIN SCORES[a,b]			
Pain score intercept	1.470	<0.0001	1.364 to1.577
Geriatric age[e]	-0.185	0.047	-0.337 to -0.003
Depression[f]	0.228	0.001	0.102 to 0.393
Time (days)[g]	-0.677	<0.0001	-0.768 to -0.587
Interaction between depression and time	0.133	0.045	0.003 to 0.263
FIXED EFFECTS FOR LOS (days)[c,d]			
LOS intercept	0.247	<0.0001	0.171 to 0.322
Geriatric age[e]	0.220	0.001	0.093 to 0.346
Depression[f]	0.089	0.096	-0.016 to 0.193

Abbreviation: CI = 95% confidence interval, LOS = length of stay.

[a]A random effects linear model of the transformed maximum pain scores was fitted, simultaneously with an accelerated failure-time lognormal model of hospital LOS postsurgery.[14] This joint mixed model accounted for the correlation between LOS and the evolution of pain scores after surgery. The distribution of the original pain scores was highly skewed with higher frequencies for severe pain scores. Maximum pain scores were previously transformed as $\log((\text{Pain Score}_{ij}+1)/(11-\text{Pain Score}_{ij}))$, where Pain Score_{ij} is the maximum daily pain score for patient i at day Time_{ij}. After this transformation, the model fitted well according to residual and random effects analyses. The pain model included a random intercept and a random slope for time and had the form $\text{Transformed Pain Scores}_{ij} = \beta_0 + \beta_1 \times \text{Geriatric Age}_i + \beta_2 \times \text{Depression}_i + \beta_3 \times \text{Time}_{ij} + \beta_4 \times \text{Depression}_i \times \text{Time}_{ij} + \alpha_{0i} + \alpha_{1i} \times \text{Time}_{ij} + e_{ij}$, where e_{ij} indicates the residuals for the pain score model for patient i at occasion j which has mean 0 and residual variance σ_e^2. The parameters β_k, $k = 1,\ldots,4$, are population-average effects (the fixed effects), whereas α_{0i} and α_{1i} are parameters specific to patient i denoting deviations from the corresponding population-averages (the random effects). The joint mixed model was fitted using the "jmre1" Stata command (StataCorp LLC, College Station, TX)[14]

[b]The variances of the random effects were 0.1384 for the pain score intercept, and 0.0916 for the time slope. The residual variance for the pain model was 0.3835.

[c]The model of LOS for patient i had the form $\log(\text{LOS}_i) = \beta_0^d + \beta_1^d \times \text{Geriatric Age}_i + \beta_2^d \times \text{Depression}_i + e_i^d$, where e_i^d is a random residual following a normal distribution with mean 0.

[d]The variance of the LOS intercept was 0.2281.

[e]The dichotomous covariate geriatric age was defined as 1 if the age of the subject was >65, and 0 otherwise.

[f]The dichotomous covariate depression was defined as 1 if the patient had a record of depression diagnosis or was under antidepressants, and 0 otherwise.

[g]Time was defined as days post spinal fusion surgery.

Impact of depression and age on individual benefits of postoperative pain management

Although treatment benefits tended to increase over time for all four groups of patients, the amount and rate of change of achieved benefits varied across groups (Table 4). For instance, at day 1, in non-geriatric patients without depression the median decrease in disease severity was 18.6% probability units compared to 5.6% in non-geriatric patients with depression

Table 4. Sample medians (and first and third quartiles) of individual benefits (x100) of postoperative pain management on days 1 through 5 for 330 patients after spinal fusion[a]

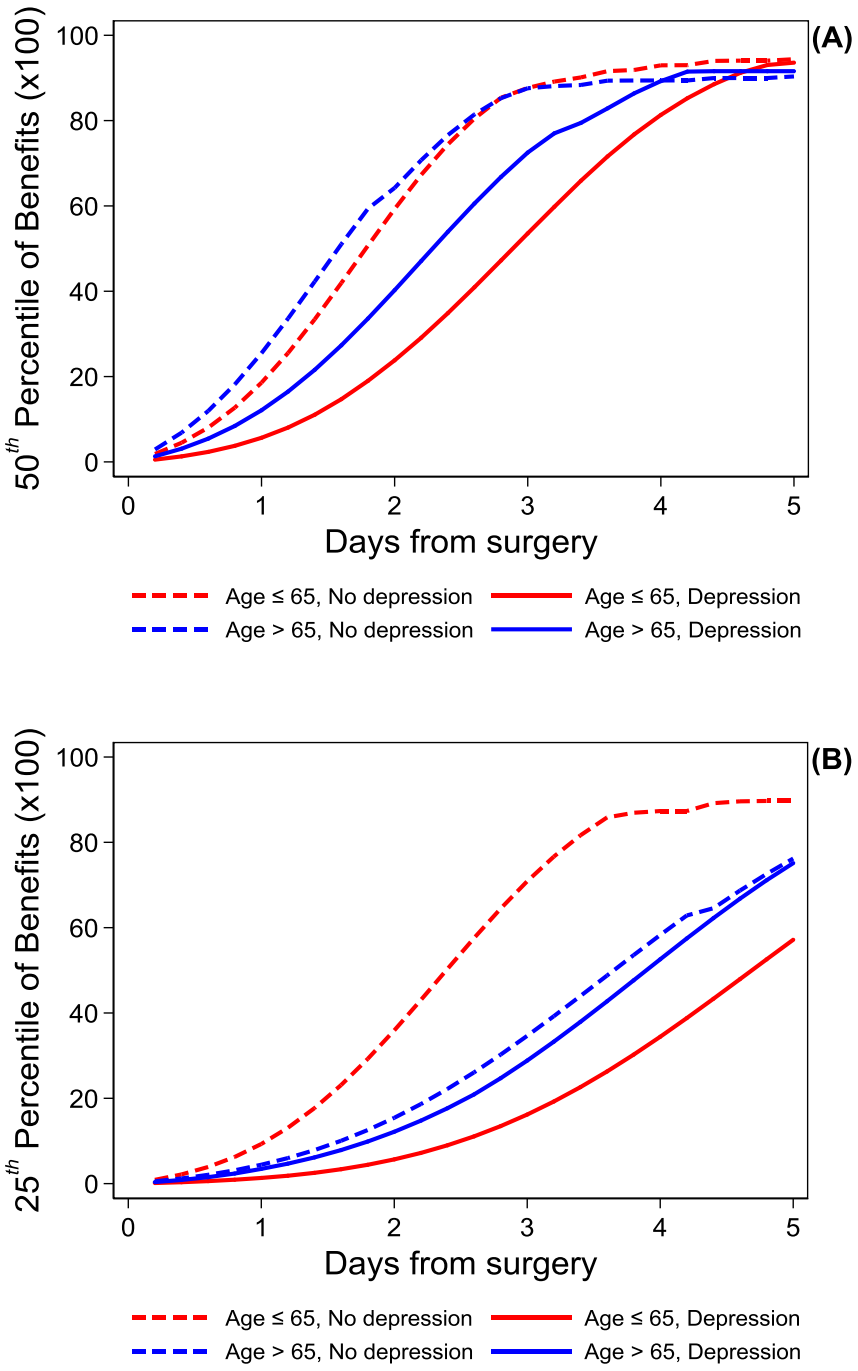
Study group	Day 1	Day 2	Day 3	Day 4	Day 5
Age > 65 and No depression (N=49)	25.5 (4.4, 34.2)	64.2 (15.4, 76.3)	87.6 (34.6, 89.4)	89.4 (58.3, 93.3)	90.4 (76.2, 93.8)
Age ≤ 65 and No depression (N=191)	18.6 (9.3, 37.1)	59.3 (35.9, 81.1)	87.7 (70.9, 91.8)	93.0 (87.3, 95.5)	94.4 (89.7, 97.0)
Age > 65 and depression (N=10)	12.1 (3.5, 19.1)	40.3 (12.2, 56.4)	72.5 (28.9, 83.5)	89.3 (52.6, 92.8)	91.6 (75.1, 96.1)
Age ≤ 65 and depression (N=80)	5.6 (1.3, 17.3)	23.8 (5.7, 56.1)	53.6 (16.2, 87.3)	81.3 (34.4, 95.2)	93.6 (57.2, 96.9)

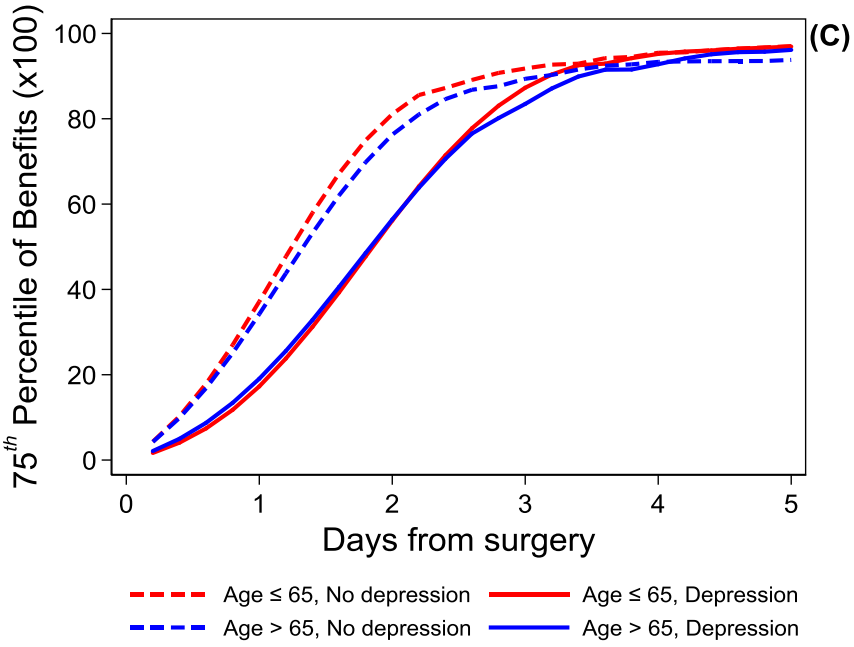
[a]The individual benefit is the increase in the probability of being in the treatment target from baseline.[7,8] The treatment target was defined as a maximum daily pain score ≤6, which corresponds to a transformed pain score ≤0.3365.

[b]The individual benefit for subject i at time t was calculated as $\hat{b}_i(t) = \left\{\Phi\left(\frac{0.3365-\hat{y}_i(t)}{\hat{\sigma}_e}\right) - \Phi\left(\frac{0.3365-\hat{y}_i(0)}{\hat{\sigma}_e}\right)\right\} \times 100,$ with $\hat{y}_i(t) = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Geriatric Age}_i + \hat{\beta}_2 \times \text{Depression}_i + \hat{\beta}_3 \times t + \hat{\beta}_4 \times \text{Depression}_i \times t + \hat{\alpha}_{0i} + \hat{\alpha}_{1i} \times t,$ where $\hat{y}_i(t)$ is the patient's predicted transformed pain score at time t; $\hat{\beta}_j$ is the maximum likelihood estimator of β_j for $j = 1, \ldots, 4$; $\hat{\alpha}_{0i}$ and $\hat{\alpha}_{1i}$ are the empirical Bayes predictors of α_{0i} and α_{1i}, respectively; Φ is the standard normal cumulative distribution function; and $\hat{\sigma}_e$ is the estimated standard deviation of the pain model residuals e_{ij}.The empirical Bayes predictors of the random effects of each patient were calculated using the "predict" command of the "jmre1" Stata command (StatCorp LLC, College Station, TX).

By day 5, the median achieved benefits were comparable for patients with or without depression. However, the first quartiles for patients with depression tended to be smaller than those for non-depressive patients of comparable age group at specific times, indicating that there were more patients with depression receiving small benefits than patients without depression after controlling for age and time.

Figure 1. Predicted evolution of individual pain management benefits (x100) after a spinal fusion surgery over days 1 through 5 for 330 patients. For a particular patient, predicted individual treatment benefits were obtained by combining the patient's data with the parameter estimates in Table 2. (A), (B), and (C) show 50th, 25th and 75th percentiles of individual benefits, respectively.





For patients with depression, non-geriatric age was associated with slower individual benefits development. For instance, in geriatric patients with depression, the median decrease in disease severity was 12.1% probability units at day 1 compared to 5.6% for non-geriatric patients with depression. On day 5, the first quartile for geriatric patients with depression (75.1%) was higher than that for non-geriatric patients with depression (57.2%). Figure 1A illustrates that for average patients with depression non-geriatric age was associated with smaller benefits, compared to geriatric age. In general, average patients with depression had much smaller benefits after controlling for age.

In patients receiving the poorest benefits from pain management the combination of depression and non-geriatric age was associated with the slowest responses whereas non-geriatric age without depression was associated with the fastest responses (Figure 1B). Interestingly, for the patients achieving the greatest benefits (Figure 1C), individual benefits were more clearly affected by depression comorbidity than by age.

Figure 2 suggests that preoperative depression diagnosis was associated with slower pain reduction after controlling for age and time. The number of non-geriatric patients who received substantial benefits on a given day post-surgery was higher for the group without than for the group with depression (Figure 2,

left panels). Even by day 5 post-surgery, there was still a much higher number of patients in the group with depression who only received minimal benefits from pain management. Similar patterns were seen in geriatric patients (Figure 3, right panels).

Effect of antidepressants on individual benefits of postoperative pain management in patients with depression

To assess whether treatment with antidepressants influenced response to pain management in patients with depression, we fitted an additional joint mixed model using only patients with depression, similar to the model in Table 3 except that the depression variable was replaced by antidepressant use. Antidepressant use was not significantly associated with baseline pain scores (p=0.283) and did not significantly modify postoperative pain reduction (p=0.53). There were no significant differences in hospital LOS (p=0.792) after surgery between patients with and without antidepressant use. Furthermore, geriatric age was not significantly associated with baseline pain scores (p=0.099) or LOS (p=0.126).

Individual benefits one day after surgery as predictors of LOS

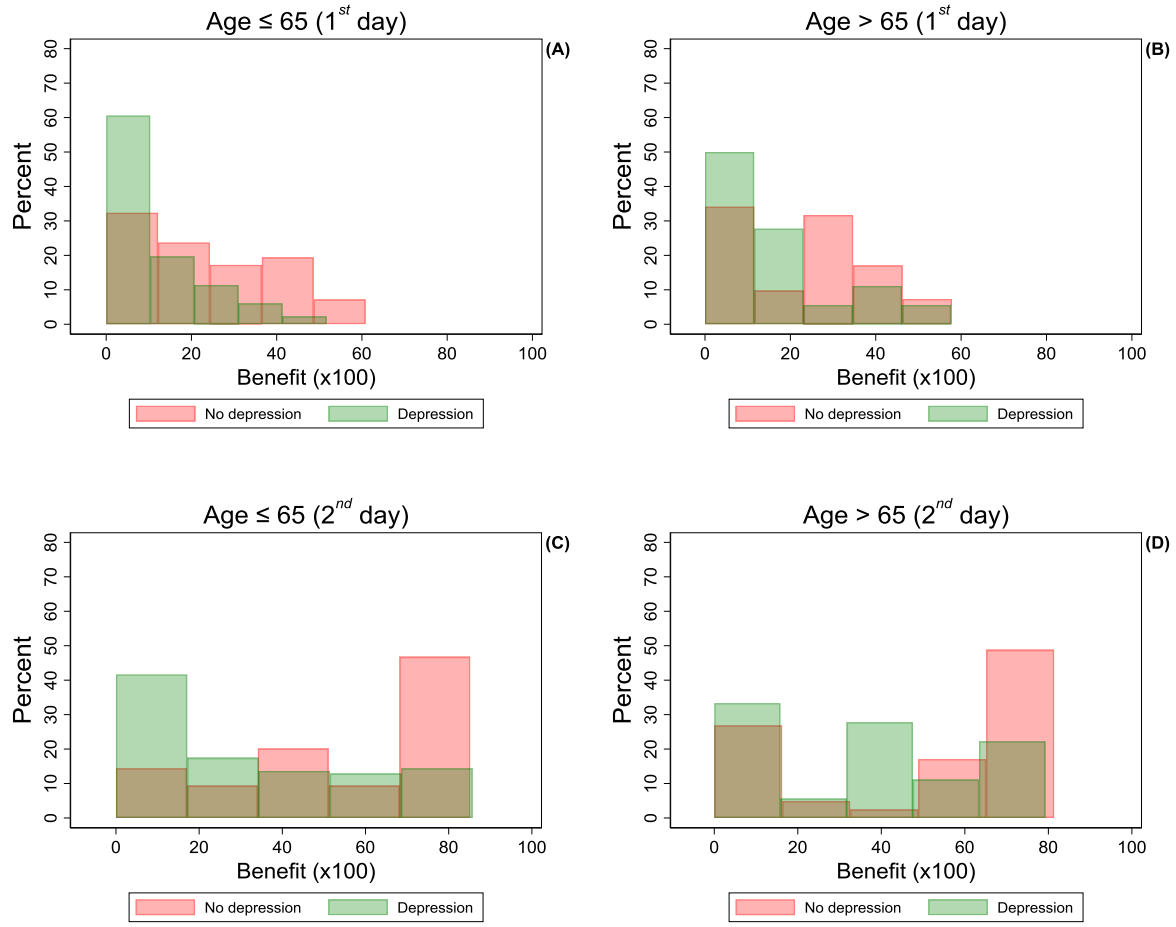
To examine whether levels of individual benefits from post-surgery pain management achieved after 1 day are predictive of hospital LOS, we compared the LOS from patients whose individual benefits were between the 1st, 2nd, 3rd, and 4th quartiles (Table 5**).** Patients with higher immediate benefits tended to have shorter LOS.

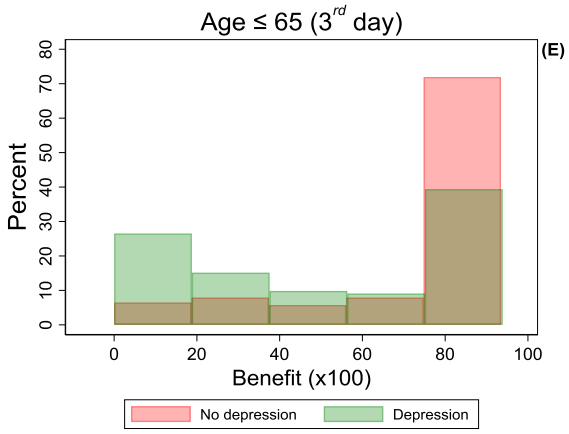
Table 5. Hospital LOS (in days) for study patients grouped by quartiles of individual pain management benefits at day 1 post spinal fusion surgery

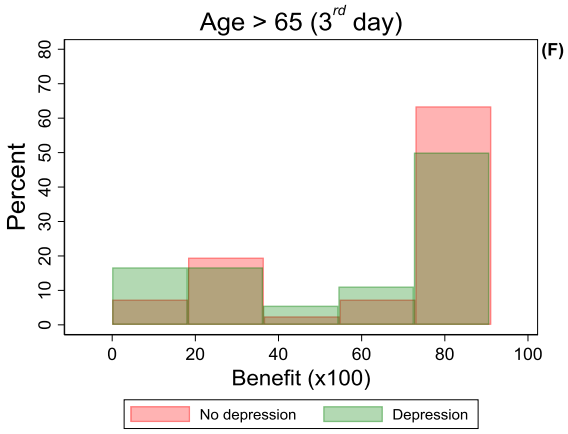
Individual Benefits	N	Mean (SD)	Median	Minimum	Maximum
1st quartile (0 to 3.28%)	82	2.44 (1.25)	2	1	5
2nd quartile (3.29 to 13.64%)	83	1.69 (0.96)	1	1	5
3rd quartile (13.65 to 27.92%)	84	1.19 (0.50)	1	1	3
4th quartile (≥27.93%)	81	1.15 (0.45)	1	1	3

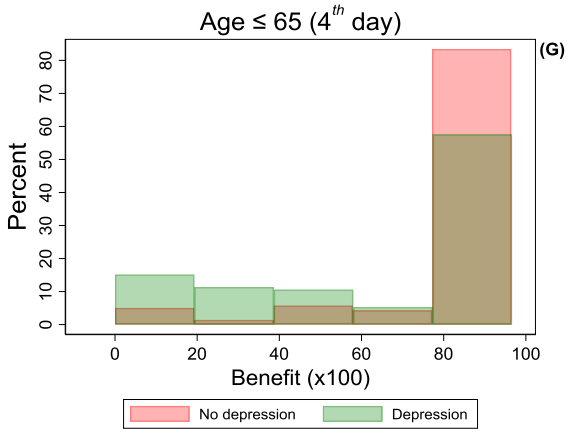
Abbreviation: LOS = length of stay.

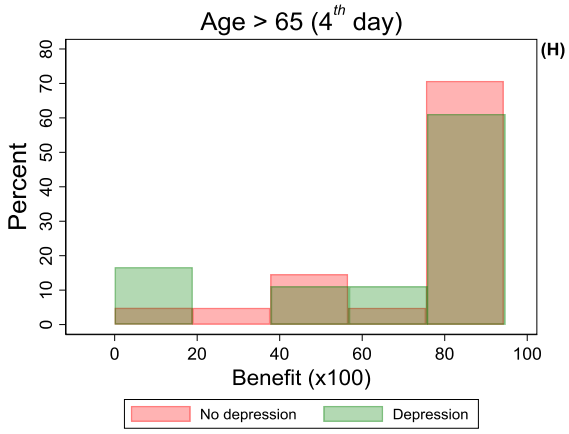
Figure 2. Histograms of predicted individual pain management benefits (x100) after a spinal fusion surgery on days 1 through 5 in patients with age ≤ 65 with or without depression (left panels) and age > 65 with or without depression (right panels), assuming the model in Table 2. (A), (C), (E), (G), and (I) are benefits for patients with age ≤ 65 with or without depression on days 1, 2, 3, 4, and 5, respectively. (B), (D), (F), (H), and (J) are benefits for patients with age > 65 with or without depression on days 1, 2, 3, 4, and 5, respectively.

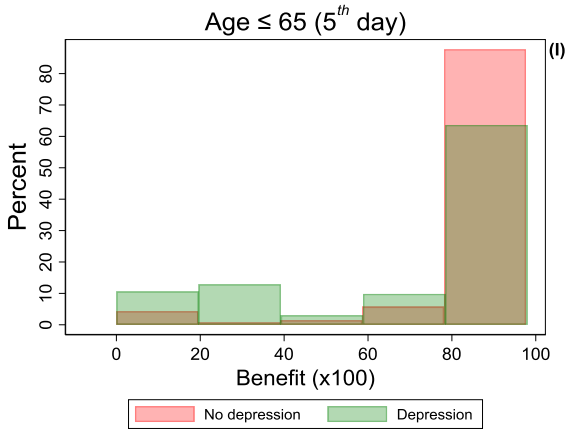


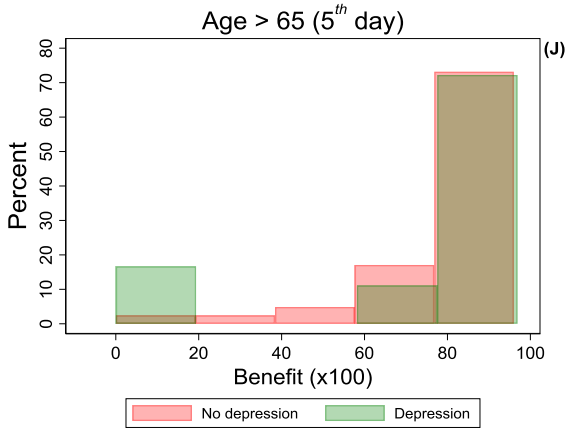

Age ≤ 65 (3rd day)
(E)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression


Age > 65 (3rd day)
(F)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression


Age ≤ 65 (4th day)
(G)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression


Age > 65 (4th day)
(H)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression


Age ≤ 65 (5th day)
(I)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression


Age > 65 (5th day)
(J)
Percent
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100
Benefit (x100)
No depression
Depression

3.4. DISCUSSION

Strengths of our statistical model

Unlike randomized clinical trials, EHR data are longitudinally unbalanced due to incomplete follow-up. This type of data is likely to have non-ignorable missingness (Touloumi, 1999; Pantazis and Touloumi, 2010) caused by termination of pain measurements due to discharge. The simultaneous modeling of LOS and pain took into consideration the correlations between them. It reduced the bias associated with unbalanced data and provided more accurate estimation of the effects of age and depression on pain scores.

Another novelty of this study is the assessment of the impact of preoperative depression and age on the individual benefits of post-operative pain management instead of focusing only on average effects. These analyses are more consistent with the goals of personalized medicine (Diaz, 2016 and 2019).

Limitations

Our study included patients with severe pain at baseline and at least one pain score and who stayed at least 1 day in the hospital. These criteria may have excluded less severe cases so our results cannot be extrapolated to them. Moreover, to increase homogeneity we selected our sample from the hospital with more cases in the EHR database. There is no way of knowing how representative this hospital sample was, although this is a typical limitation of observational data.

The pain scores used in this study were self-reported values. Patient-reported measures such as pain scores and levels of satisfaction are important measures for evaluating treatment effects (Lotzke, 2016). They could be biased, however, since each patient may have different levels of sensitivity and expectation. However, predictors of individual benefits, which are on a probability scale, compare baseline pain severities with post-treatment severities within a patient, canceling out potential individual biases in the perception of pain.

Comparison with prior studies

In the present study, almost half (46%) the patients undergoing spinal fusion surgery had depression. Pain scores decreased at a slower pace after surgery in patients with depression. The effect of geriatric age was not as dramatic but did have an impact on the individual benefits of pain management in

subsets of patients. For example, we could see the effects of geriatric age in those who were not doing so well and those with average benefits, although geriatric age did not show substantial effects on the benefits for the patients who were responding well. We also found associations between higher baseline pain scores, longer LOS, and slower speeds of postoperative pain relief. Our study confirms the finding from earlier studies that a high proportion of patients with chronic pain have depression (Greden, 2009). It has been demonstrated that depression and chronic pain go together, making it hard to determine cause and effect (Gaudin, 2017; Trivedi, 2004; Anderson et al., 2015; Greden, 2009; Arnold et al., 2012). Several studies showed that depression and age both have an impact on the feeling of pain (Gaudin, 2017; Trivedi, 2004; Anderson et al., 2015; Gerbershagen et al., 2014; Weinmann et al., 2017). Geriatric patients tend to report lower pain levels, which could be due to their decreased sensitivity to pain (Gerbershagen et al., 2014; Weinmann et al., 2017).

An earlier study found that females tended to have slightly higher postoperative pain levels as compared to males (Gerbershagen et al., 2014). In our study, however, gender was not significant in a joint mixed model that adjusted for depression. More females (35%) had a preoperative depression diagnosis than males (19%), and the p-value for gender before adjusting for depression was smaller although still not significant ($p= 0.178$). Thus, it is possible that depression mediated the previously reported relationship between female gender and pain to some degree.

Antidepressants had no significant effects

To rule out the possibility that antidepressant medication explains the observed slower response to pain management in patients with depression, we analyzed the effects of antidepressant treatments in patients with depression. We found that antidepressants were not significantly associated with baseline pain scores or the response to pain management. Thus, the slow response to pain management in patients with depression may be due to the comorbidity itself instead of antidepressant medication.

Individual benefits

This study compared individual benefits of pain management among four groups of patients determined by depression diagnosis and age. An examination of median benefits was not enough, and other

subgroups of individuals emerged (Figure 1). In "average" patients, age played an important role in those with depression, who were prone to receive less benefit (Figure 1A). In contrast, among patients tending to receive the smallest benefits (Figure 1B), younger patients without depression achieved some benefit quicker than geriatric patients with depression, whereas younger patients with depression were the least benefitted from pain management. Moreover, among patients achieving the highest benefits (Figure 1C), the effect of age on treatment benefits was negligible compared to the effect of depression. Our finding that the effect of age is unimportant in patients receiving high benefits is consistent with the results of a previous study that found that, although elderly patients reported lower pain scores post total knee replacement, their functional impairment caused by pain did not differ from younger patients.[6]

LOS

Interestingly, patients who received less benefit from one day of post-surgery pain management tended to stay longer at the hospital (Table 5), suggesting that early benefit measurements may serve as predictors of hospital LOS after surgery.

Conclusion

Our study revealed that preoperative depression and geriatric age are important factors affecting individual benefits of postoperative pain management in patients undergoing spinal fusion surgery. Depression had a negative impact on pain relief, while age had varied effects depending on depression status and potentially other traits. Moreover, joint mixed models are useful tools for analyzing unbalanced longitudinal EHR data caused by hospital lengths of stay that are related to treatment response. Finally, individual benefit predictions provided a practical way to evaluate the performance of postoperative pain management.

3.5. References

Adogwa O., Parker S.L., Shau D.N., et al. (2013) Preoperative Zung depression scale predicts patient satisfaction independent of the extent of improvement after revision lumbar surgery. Spine J. 13(5):501-506.

Albers D.J., Elhadad N., Claassen J., Perotte R., Goldstein A., Hripcsak G. (2018) Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms. J Biomed Inform. 78(2):87–101.

Anderson J.T., Haas A.R., Percy R., Woods S.T., Ahn U.M., Ahn N.U. (2015) Clinical depression is a strong predictor of poor lumbar fusion outcomes among workers' compensation subjects. Spine. 40(10):748-756.

Arnold L.M., Palmer R.H., Gendreau R.M., Chen W. (2012) Relationships among pain, depressed mood, and global status in fibromyalgia patients: post hoc analyses of a randomized, placebo-controlled trial of milnacipran. Psychosomatics 53(4):371-379.

Diaz F.J. (2016) Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat Med. 35(23):4077–4092.

Diaz F.J. (2019) Estimating individual benefits of medical or behavioral treatments in severely ill patients. Stat Methods Med Res. 28(3):911-927.

Gaudin D., Krafcik B.M., Mansour T.R., Alnemari A. (2017) Considerations in spinal fusion surgery for chronic lumbar pain: psychosocial factors, rating scales, and perioperative patient education-a review of the literature. World Neurosurg. 2017;98(2):21-27.

Gerbershagen H.J., Pogatzki-Zahn E., Aduckathil S., et al. (2014) Procedure-specific risk factor analysis for the development of severe postoperative pain. Anesthesiology. 120(5):1237-1245.

Greden J.F. (2009) Treating depression and pain. J Clin Psychiatry. 70(6):e16.

Ibrahim J.G., Molenberghs G. (2009) Missing data methods in longitudinal studies: a review. Test (Madr). 18(1):1-43.

Lotzke H., Jakobsson M., Brisby H., et al. (2016) Use of the PREPARE (PREhabilitation, Physical Activity and exeRcisE) program to improve outcomes after lumbar fusion surgery for severe low back pain: a study protocol of a person-centred randomised controlled trial. BMC Musculoskelet Disord. 17(1):349.

Pantazis N., Touloumi G. (2010) Analyzing longitudinal data in the presence of informative drop-out: The jmre1 command. Stata J. 10(2):226-251.

Shaw A.D., Mythen M.G., Shook D., et al. (2018) Pulmonary artery catheter use in adult patients undergoing cardiac surgery: a retrospective, cohort study. Perioper Med (Lond). 7(24)

Touloumi G., Pocock S.J., Babiker A.G., Darbyshire J.H. (1999) Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. Stat Med. 18(10):1215-1233.

Trivedi M.H. (2004) The link between depression and physical symptoms. Prim Care Companion J Clin Psychiatry. 2004;6(Suppl 1):12-16.

Urman R.D., Boing E.A., Pham A.T., et al. (2018) Improved outcomes associated with the use of intravenous acetaminophen for management of acute post-surgical pain in cesarean sections and hysterectomies. J Clin Med Res. 10(6):499-507.

Weinmann C., Komann M., Meissner W. (2017) Tough cookies: the older the patients, the more pain tolerating? EJA. 34(55):215.

3.6. Appendix: Stata Code

3.6.1. Stata Code for Chapter One

evaluate_disorganized.do.

(Stata do file. Run simulations from this file.)

```
trueparam_disorganized, delta(0.5) canna01(0) predorig(3) h(1) numpat(1000) reseed(24)
display "canna01: "$canna01
display "Prediction origin: t="$PredOrig
display "Prediction horizon: h="$h
display "delta= " $delta
display "Median of relative biases: " r(MedianRelBias)
display "Minimum of relative biases: " r(MinRelBias)
display "Maximum of relative biases: " r(MaxRelBias)
display "Median of correlations between predicted and true transformed benefits:  " r(MedianCorr)
display  "Minimum of correlations: " r(MinCorr)
display "Maximum of correlations:  " r(MaxCorr)
```

trueparam_disorganized.ado.

(Stata ado program that performs the Monte Carlo simulations. This program is called by evaluate_disorganized.do)

```
program trueparam_disorganized, rclass
version 15.1
syntax, delta(numlist max=1 >=0)  canna01(numlist integer >=0 <=1) predorig(integer) h(integer) ///
                [numpat(integer 1000) reseed(integer -1 )]
clear
*Seed for simulating random effects with drawnorm command
if `reseed'<0 {
global reseed " "
```

```
}
else {
global reseed "seed(`reseed')"
}
**********
global canna01=`canna01'  // enter 1 if patient used cannabis
**********
global PredOrig=`predorig' //PredOrig is the prediction origin (a time point).
global h=`h'  //h=horizon; enter a negative number or 0 for retrospective measurement of benefits;
**********
global NumPat=`numpat'   //Enter number of simulated patients
**********
global delta=`delta'   // Enter distance from a true parameter to parameter estimate in Table 1 in standard
error units
**********
* 64 is the total number of possible combinations of true parameter values for a fixed value of delta.
* There are 6 model parameters and, therefore,  2^ 6 = 64
set matsize 64
if $NumPat>64 {
if $NumPat<=11000 set matsize $NumPat
else {
display as error "Number of simulated patients cannot be higher than 11000"
exit, clear
}
}
```

TrueParam_Module1_disorganized

```
TrueParam_Module2_disorganized
clear
quietly svmat Results, names(col)    // convert the matrix into a data set
save
"Results_delta${delta}_canna01${canna01}_PredOrig${PredOrig}_h${h}_NumPat${NumPat}.dta",
replace
quietly summarize MeanBias,detail
return scalar MedianMeanBias=r(p50)
return scalar MinMeanBias=r(min)
return scalar MaxMeanBias=r(max)
quietly summarize SDBias, detail
return scalar MedianSDBias=r(p50)
return scalar MinSDBias=r(min)
return scalar MaxSDBias=r(max)
quietly summarize RelBias, detail
return scalar MedianRelBias=r(p50)
return scalar MinRelBias=r(min)
return scalar MaxRelBias=r(max)
quietly summarize Correlation, detail
return scalar MedianCorr=r(p50)
return scalar MinCorr=r(min)
return scalar MaxCorr=r(max)
quietly summarize MeanTrueBenef, detail
return scalar MedianMeanTrueBenef=r(p50)
return scalar MinMeanTrueBenef=r(min)
return scalar MaxMeanTrueBenef=r(max)
```

```
**********
end
```

TrueParam_Module1_disorganized.ado.

(Stata ado program used by trueparam_disorganized.)

```
set more off
set matsize 11000
*This reads the estimates of the model reported in Table 1 of paper
estimates use  "Fitted_model"
*This gets the variance covariance matrix of estimates
matrix VCe=e(V)
**********************************************
*Vector of estimated fixed effects from Model in Table 1 is created
global b1=_b[dis_lt4:_cons]
global b2=_b[dis_lt4:canna01]
global b3=_b[dis_lt4:pt1]
global b4=_b[dis_lt4:pt2]
global b5=_b[dis_lt4:pt3]
matrix bGLS=($b1   \   ///
                    $b2   \   ///
                    $b3   \   ///
                     $b4   \   ///
                     $b5)
*The variance-covariance matrix D of random effects from model in Table 1 is created
global D11=_b[/var(_cons[id])]
matrix D=$D11
end
```

TrueParam_Module2_disorganized.ado.

(Stata ado program used by trueparam_disorganized.)

```
if $delta!=0 {
   matrix Results=J(64,5,0)   //2^6=64
   local deltalist -$delta  $delta
   }
else {
   matrix Results=J(1,5,0)
   local deltalist 0
    }
**********************
matrix colnames Results = MeanBias SDBias RelBias Correlation MeanTrueBenef
local RowOfResults=1
foreach delta1 of numlist `deltalist' {
foreach delta2 of numlist `deltalist' {
foreach delta3 of numlist `deltalist' {
foreach delta4 of numlist `deltalist' {
foreach delta5 of numlist `deltalist' {
foreach delta6 of numlist `deltalist' {
*True fixed effects are computed
global b1True=$b1 + `delta1'*sqrt(VCe[5,5])
global b2True=$b2 + `delta2'*sqrt(VCe[1,1])
global b3True=$b3 + `delta3'*sqrt(VCe[2,2])
global b4True=$b4 + `delta4'*sqrt(VCe[3,3])
global b5True=$b5 + `delta5'*sqrt(VCe[4,4])
matrix bTrue=($b1True   \   ///
```

```
                $b2True   \   ///
                $b3True   \   ///
                $b4True   \   ///
                $b5True)
matrix list bTrue
*True variance covariance matrix is computed
global D11True=$D11+`delta6'*sqrt(VCe[6,6])   /* variance of intercept */
matrix DTrue=$D11True
****************************************
display "Simulation `RowOfResults' for canna01=$canna01, Delta=$delta, Prediction Origin=$PredOrig,
Horizon=$h"
clear
TrueParam_Module3_disorganized
matrix Results[`RowOfResults',1]=$MeanBias
matrix Results[`RowOfResults',2]=$SDBias
matrix Results[`RowOfResults',3]=$RelBias
matrix Results[`RowOfResults',4]=$Correlation
matrix Results[`RowOfResults',5]=$MeanTrueBenef
local RowOfResults=`RowOfResults'+1
}
}
}
}
}
End
```

TrueParam_Module3_disorganized.ado.

(Stata ado program used by trueparam_disorganized.)

```
*Design matrix Z for random effects
matrix A=J(7, 1, 1)
matrix pt1=J(7, 1, 0)
forvalues i=1/7{
        matrix pt1[`i', 1]=P[1,1]*(`i'-1) + P[1,2]*(`i'-1)^2 + P[1,3]*(`i'-1)^3 + P[1,4]
}
matrix Z=A
matrix colnames Z=Intercept
**********************************************
*Design matrix X for fixed effects
matrix A=J(7, 1, 1)
matrix B=J(7, 1, $canna01)
matrix pt2=J(7, 1, 0)
forvalues i=1/7{
        matrix pt2[`i', 1]=P[2,1]*(`i'-1) + P[2,2]*(`i'-1)^2 + P[2,3]*(`i'-1)^3 + P[2,4]
}
matrix pt3=J(7, 1, 0)
forvalues i=1/7{
        matrix pt3[`i', 1]=P[3,1]*(`i'-1) + P[3,2]*(`i'-1)^2 + P[3,3]*(`i'-1)^3 + P[3,4]
}
matrix X=A,B,pt1,pt2,pt3
matrix colnames X=Intercept canna01 ptime1 ptime2 ptime3
**********************************************
```

```
if $PredOrig==0|$PredOrig==1|$PredOrig==2|$PredOrig==3|$PredOrig==4|$PredOrig==5
|$PredOrig==6 {


 if $PredOrig==0 {
 matrix Z=Z[1..1,1...]   //
 matrix X=X[1..1,1...]
  }
 if $PredOrig==1{
 matrix Z=Z[1..2,1...]
 matrix X=X[1..2,1...]
 }
 if $PredOrig==2{
 matrix Z=Z[1..3,1...]
 matrix X=X[1..3,1...]
 }
 if $PredOrig==3{
 matrix Z=Z[1..4,1...]
 matrix X=X[1..4,1...]
 }
 if $PredOrig==4{
 matrix Z=Z[1..5,1...]
 matrix X=X[1..5,1...]
  }
   if $PredOrig==5{
 matrix Z=Z[1..6,1...]
 matrix X=X[1..6,1...]
```

```
 }


     if $PredOrig==6{
 matrix Z=Z[1..7,1...]
 matrix X=X[1..7,1...]
 }
}
else{
display as error "Values for predOrig should be 0,1,2,3,4,5,or 6"
exit, clear
}
*********************************************
if !(1<=$PredOrig+$h & $PredOrig+$h<=6) {
display as error "predorig+h should be in the interval [1,6]"
exit, clear
}
*********************************************
*First we simulate the patients
*The simulated random intercept LambdaR has mean zero
clear
quietly drawnorm  LambdaR, n($NumPat)  cov(DTrue)  $reseed  // DTrue was created in Module2
mkmat LambdaR, matrix(RanEff)  // Each row of matrix RanEff corresponds to a set of random effects (for
intercept and ptime1) for one simulated patient
*Columns of matrix MatbTrue will contain the true fixed effects repeatedly
matrix MatbTrue=bTrue
forvalues i=2/$NumPat {
```

```
        matrix MatbTrue=MatbTrue,bTrue    // each column has a vector of bTrue
}
*Calculate linear predictor
matrix RanEfft=RanEff'
matrix ZR=Z*RanEfft
matrix XB=X*MatbTrue
matrix ata=ZR+XB
*For each element of anta, calculate p. This will be the predicted P, which can be used to calculate the predicted benefit as well as to simulate y.
matrix p=J(rowsof(ata), colsof(ata), 0)  // number of time points by number of patients
matrix y=J(rowsof(ata), colsof(ata), 0)
/* begin loop */
local i=1           // i for time
while `i'<=rowsof(ata){
  local j=1         // j for patients
  while `j'<=colsof(ata){
    matrix p[`i',`j']=exp(ata[`i',`j'])/(1+exp(ata[`i',`j']))

        matrix y[`i',`j']=rbinomial(1, p[`i',`j'])

    local j = `j' + 1
  }
   local i = `i' + 1
}
/* end loop */
*Create dataset xy with y and Xs for 1000 patients in rows by appending matrices. Include patient id
```

```
matrix X1=X
forvalues i=2/$NumPat {
        matrix X1=X1\X
}
matrix y1=y[1..., 1]
forvalues j=2/$NumPat {
        matrix y2=y[1..., `j']
        matrix y1=y1\y2
}
matrix id=J(rowsof(ata), 1, 1)
forvalues j=2/$NumPat {
        matrix id2=J(rowsof(ata), 1, `j')
        matrix id=id\id2
}
matrix xy=y1, X1, id
matrix colnames xy=dis_lt4 Intercept canna01 pt1 pt2 pt3 id
// convert to dataset
clear
svmat xy, names(col)
predict PrRanEff* , reffects
duplicates drop id, force
keep PrRanEff* id
svmat RanEff, names(col)
spearman PrRanEff1 LambdaR
*******************************************************************************
*Compute predicted benefit and true benefit for each patient
```

```
quietly generate ppred2=1/(1+ exp(-($b1 + PrRanEff1 + $b2*$canna01  + $b3 * pt1[$PredOrig+$h+1, 1]
+ ///                                $b4 * pt2[$PredOrig+$h+1, 1] + $b5 * pt3[$PredOrig+$h+1, 1] )))
quietly generate ppred1=1/(1+ exp(-($b1 + PrRanEff1 + $b2*$canna01 + $b3* pt1[1, 1] + ///
                                                $b4 * pt2[1, 1] +  $b5 * pt3[1, 1] )))
quietly generate PredBenef=ppred2-ppred1
quietly generate ptrue2=1/(1+ exp(-($b1True + LambdaR + $b2True*$canna01  + $b3True *
pt1[$PredOrig+$h+1, 1] + ///
                                                $b4True * pt2[$PredOrig+$h+1, 1] + $b5True *
pt3[$PredOrig+$h+1, 1] )))
quietly generate ptrue1=1/(1+ exp(-($b1True + LambdaR + $b2True*$canna01  + $b3True * pt1[1, 1] + ///
                                                $b4True * pt2[1, 1] + $b5True * pt3[1, 1] )))
quietly generate TrueBenef=ptrue2-ptrue1
keep PredBenef TrueBenef
*************************************************************************************
*Individual bias is computed
quietly generate double Bias=PredBenef-TrueBenef
*Individual relative bias is computed
quietly summarize Bias, detail
global MeanBias=r(mean)  //Mean bias
global SDBias=sqrt(r(Var))  // SD of bias
quietly summarize TrueBenef, detail
global MeanTrueBenef=r(mean)
global RelBias=($MeanBias/$MeanTrueBenef)*100  // Relative bias
*************************************************************************************
*Computation of correlation between predicted and true benefit
```

```
quietly spearman TrueBenef PredBenef
global Correlation=r(rho)
end
```

3.6.2. Stata Code for Chapter Two

evaluate.do.

(Stata do file. Run simulations from this file.)

```
trueparam, delta(0) age(0) depress(0) predorig(5) h(0) numpat(1000) reseed(24) erseed(30)
display "age: "$age
display "depress: "$depress
display "Prediction origin: t="$PredOrig
display "Prediction horizon: h="$h
display "delta= " $delta
display "Median of relative biases: " r(MedianRelBias)
display "Minimum of relative biases: " r(MinRelBias)
display "Maximum of relative biases: " r(MaxRelBias)
display "Median of correlations between predicted and true transformed benefits:  " r(MedianCorr)
display  "Minimum of correlations: " r(MinCorr)
display "Maximum of correlations:  " r(MaxCorr)
```

trueparam.ado.

(Stata ado program that performs the Monte Carlo simulations. This program is called by evaluate.do.)

```
program trueparam, rclass
version 15.1
syntax, delta(numlist max=1 >=0) age(numlist integer >=0 <=1) depress(numlist integer >=0 <=1)
predorig(integer) h(integer) ///
                    [numpat(integer 1000) reseed(integer -1 ) erseed(integer -1)]
clear
```

```
if `reseed'>=0&`erseed'>=0&`reseed'==`erseed'{
display as error "reseed has to be different from erseed."
exit, clear
}
*Seed for simulating random effects with drawnorm command
if `reseed'<0 {
global reseed " "
}
else {
global reseed "seed(`reseed')"
}
*Seed for simulating model error terms
if `erseed'<0 {
global erseed " "
}
else {
global erseed "seed(`erseed')"
}
**********
global age=`age'                    // enter 1 if patient's age >65, 0 otherwise
global depress=`depress'            // enter 1 if patient had depression, 0 otherwisec
global delta=`delta'
**********
global y=0.3365                     // The treatment target is <=6; The transformation log((6+1)/(11-
6)) gives 0.3365
global PredOrig=`predorig'          //PredOrig is the prediction origin (a time point).
```

```
global h=`h'                    //h=horizon
**********************
global NumPat=`numpat'          //Enter number of simulated patients
* 4096 is the total number of possible combinations of true parameter values for a fixed value of delta.
* There are 12 model parameters that we need to calculate the true benefit and, therefore, 2^12=4096)
set matsize 4096
if $NumPat>4096 {
if $NumPat<=11000 set matsize $NumPat
else {
display as error "Number of simulated patients cannot be higher than 11000"
exit, clear
}
}
TrueParam_Module1
TrueParam_Module2
clear
quietly svmat Results, names(col)
save
"Results_delta${delta}_age${age}_depress${depress}_PredOrig${PredOrig}_h${h}_NumPat${NumPat
}.dta", replace
quietly summarize RelBias, detail
return scalar MedianRelBias=r(p50)
return scalar MinRelBias=r(min)
return scalar MaxRelBias=r(max)
quietly summarize Correlation, detail
return scalar MedianCorr=r(p50)
```

return scalar MinCorr=r(min)

return scalar MaxCorr=r(max)

****

End

TrueParam_Module1.ado.

(Stata ado program used by trueparam.)

set matsize 11000

**** This reads the estimates of the model reported in Table 1 of paper

estimates use "Fitted_model.ster"

**** Covariance matrix

matrix D=e(cov_re)

global D11 D[1, 1]

global D12 D[1, 2]

global D13 D[1, 3]

global D21 D[2, 1]

global D22 D[2, 2]

global D23 D[2, 3]

global D31 D[3, 1]

global D32 D[3, 2]

global D33 D[3, 3]

**** Extract fixed effects

matrix B=e(b)'

**** Fixed effects for pain score model

global b4=_b[Marker:_M]

global b5=_b[Marker:_Mage_gt65]

global b6=_b[Marker:_I_Mdepress_1]

```
global b7=_b[Marker:_Mtime]

global b8=_b[Marker:_I_MdX_Mtim_1]

matrix b=( $b4 \   ///
                $b5 \   ///
                $b6 \   ///
                $b7 \   ///
                $b8)

**** This gets the variance covariance matrix of fixed effects estimates

matrix VCe=e(V)                 // We will need the SE for fixed effects of the marker model in module 2.

**** Variance of the error term for the pain score model in Table 1

global VarErr=e(var_eij)

**** Obtaining standard error of determinant of principal minor of D (eliminating 3rd row and 3rdcolumn)

local D11 D[1, 1]

local D12 D[1, 2]

local D13 D[1, 3]

local D21 `D12'

local D22 D[2, 2]

local D23 D[2, 3]

local D31 `D13'

local D32 `D23'

local D33 D[3, 3]

**** Computation of principal minor of D and its SE

global pminor2=`D11'*`D22'-(`D21')^(2)

global SEpminor2=$pminor2*0.6  // We have to give a value since we cannot calculate the SE of principal
minor

**** Computation of determinant of D and its standard error
```

```
Global                  detD=`D11'*(`D22'*`D33'-(`D32')^(2))-(`D21')^(2)*`D33'+2*`D21'*`D31'*`D32'-
`D22'*(`D31')^(2)
global SEdetD=$detD*0.6
*****************************************
End
```

TrueParam_Module2.ado.

(Stata ado program used by trueparam.)

```
 if $delta!=0 {
            matrix Results=J(4096,2,0)
    local deltalist -$delta  $delta
    }
else {
            matrix Results=J(1,2,0)
    local deltalist 0
     }
***********************
matrix colnames Results = RelBias Correlation
local RowOfResults=1
foreach delta1 of numlist `deltalist' {
foreach delta2 of numlist `deltalist' {
foreach delta3 of numlist `deltalist' {
foreach delta4 of numlist `deltalist' {
foreach delta5 of numlist `deltalist' {
foreach delta6 of numlist `deltalist' {
foreach delta7 of numlist `deltalist' {
foreach delta8 of numlist `deltalist' {
```

```
foreach delta9 of numlist `deltalist' {
foreach delta10 of numlist `deltalist' {
foreach delta11 of numlist `deltalist' {
foreach delta12 of numlist `deltalist' {
**** True fixed effects are computed
global b4True=$b4 + `delta1'*sqrt(VCe[4,4])
global b5True=$b5 + `delta2'*sqrt(VCe[5,5])
global b6True=$b6 + `delta3'*sqrt(VCe[6,6])
global b7True=$b7 + `delta4'*sqrt(VCe[7,7])
global b8True=$b8 + `delta5'*sqrt(VCe[8,8])
matrix bTrue=($b4True   \   ///
                    $b5True   \   ///
                    $b6True   \   ///
                    $b7True   \   ///
                     $b8True )
**** True variance covariance matrix of the random effects is computed
global D11True=$D11+`delta6'*0.228*0.2
global D12True=$D12+`delta7'*0.0931
global D13True=$D13+`delta8'*0.0942
global D21True=$D12True
local pminor2True=$pminor2+`delta9'*$SEpminor2 //The Variance-Covariance matrix was
reparametrized to get a positive definite matrix
lobal D22True=(`pminor2True'  +($D21True)^(2))/$D11True
global D23True=$D23+`delta10'*0.2008
global D31True=$D13True
global D32True=$D23True
```

```
local detDTrue=$detD+`delta11'*$SEdetD    //The Variance-Covariance matrix of random effects was reparametrized to get a positive definite matrix
global                    D33True=((`detDTrue'+$D11True*($D32True)^(2)-(2*$D21True*$D32True-$D22True*$D31True)*$D31True  ) / ($D11True*$D22True-($D21True)^(2)))
matrix DTrue=( $D11True  , $D12True  , $D13True  \  ///
                    $D21True  , $D22True  , $D23True  \  ///
                    $D31True  , $D32True  , $D33True  )
****************************************
**** True error variance is computed
global VarErrTrue=$VarErr+`delta12'* 0.38353*0.2
*TrueParam_Module3 simulates the random effects, creates design matrices of the appropriate size (according to PredOrig and h),
*and simulates the responses of patients, which are placed in matrix Y.
*The random effects plus their corresponding fixed effects are also saved in the database temporarily.
display "Simulation `RowOfResults' for age=$age, depress=$depress,Delta=$delta, Prediction Origin=$PredOrig, Horizon=$h"
clear
TrueParam_Module3
*TrueParam_Module4 computes the BLUPs. They are saved in database.
TrueParam_Module4
*TrueParam_Module5 computes the empirical Bayesian predictors of benefits and true benefits (and save them temporarily in database)
TrueParam_Module5
matrix Results[`RowOfResults',1]=$RelBias
matrix Results[`RowOfResults',2]=$Correlation
local RowOfResults=`RowOfResults'+1
```

```
}
}
}
}
}
}
}
}
}
}
End
```

TrueParam_Module3.ado.

(Stata ado program used by trueparam.)

```
if $PredOrig==0|$PredOrig==1|$PredOrig==2|$PredOrig==3|$PredOrig==4 |$PredOrig==5 {
  if $PredOrig==0 {
  matrix Z=(0, 1, 0 \ ///
                1, 0, 0)
  matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
                              1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
  local errors "eps01"
  }
  if $PredOrig==1{
  matrix Z=(0, 1, 0 \ ///
                0, 1, 1 \ ///
```

```
                1, 0, 0)
matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
                    0, 0, 0, 1, 1*$age, 1*$depress, 1, 1*$depress \  ///
                    1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
local errors  "eps01 eps02"
}
if $PredOrig==2{
matrix Z=(0, 1, 0 \ ///
                0, 1, 1 \ ///
                0, 1, 2 \ ///
                1, 0, 0)
matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
                0, 0, 0, 1, 1*$age, 1*$depress, 1, 1*$depress \  ///
                0, 0, 0, 1, 1*$age, 1*$depress, 2, 2*$depress \  ///
                1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
local errors "eps01 eps02 eps03"
}
if $PredOrig==3{
matrix Z=(0, 1, 0 \ ///
                0, 1, 1 \ ///
                0, 1, 2 \ ///
                0, 1, 3 \ ///
                1, 0, 0)
matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
                0, 0, 0, 1, 1*$age, 1*$depress, 1, 1*$depress \  ///
                0, 0, 0, 1, 1*$age, 1*$depress, 2, 2*$depress \  ///
```

```
                0, 0, 0, 1, 1*$age, 1*$depress, 3, 3*$depress \  ///
                1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
local errors  "eps01 eps02 eps03 eps04"
}
if $PredOrig==4{
matrix Z=(0, 1, 0 \ ///
                0, 1, 1 \ ///
                0, 1, 2 \ ///
                0, 1, 3 \ ///
                0, 1, 4 \ ///
                1, 0, 0)
matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
                        0, 0, 0, 1, 1*$age, 1*$depress, 1, 1*$depress \  ///
                        0, 0, 0, 1, 1*$age, 1*$depress, 2, 2*$depress \  ///
                        0, 0, 0, 1, 1*$age, 1*$depress, 3, 3*$depress \  ///
                        0, 0, 0, 1, 1*$age, 1*$depress, 4, 4*$depress \  ///
                        1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
local errors "eps01 eps02 eps03 eps04 eps05"
}
if $PredOrig==5{
matrix Z=(0, 1, 0 \ ///
        0, 1, 1 \ ///
        0, 1, 2 \ ///
        0, 1, 3 \ ///
        0, 1, 4 \ ///
        0, 1, 5 \ ///
```

```
            1, 0, 0)
  matrix X=(0, 0, 0, 1, 1*$age, 1*$depress, 0, 0*$depress \  ///
           0, 0, 0, 1, 1*$age, 1*$depress, 1, 1*$depress \  ///
            0, 0, 0, 1, 1*$age, 1*$depress, 2, 2*$depress \  ///
            0, 0, 0, 1, 1*$age, 1*$depress, 3, 3*$depress \  ///
            0, 0, 0, 1, 1*$age, 1*$depress, 4, 4*$depress \  ///
             0, 0, 0, 1, 1*$age, 1*$depress, 5, 5*$depress \  ///
            1, 1*$age, 1*$depress, 0, 0, 0, 0, 0)
  local errors "eps01 eps02 eps03 eps04 eps05 eps06"
  }
}
else{
display as error "Values for predOrig should be 0,1,2,3,4, or 5"
exit, clear
}
matrix colnames Z=_D _M _Mtime
matrix colnames X=_D _Dage _Ddepress _M _Mage _Mdepress _Mtime _Mdepresstime
*********************************************
if !(1<=$PredOrig+$h & $PredOrig+$h<=5) {
display as error "predorig+h should be in the interval [1,5]"
exit, clear
}
*********************************************
*Variance covariance matrix of error terms
matrix RTrue=I(rowsof(Z)-1)  *  $VarErrTrue
*Simulation of pain scores Y
```

```
*First we simulate the patients


*The random effects are simulated.
*A particular value of vector (re_D re_M re_Mtime) correspond to one patient.
*The simulated random variables  re_D re_M re_Mtime have mean zero
clear
quietly drawnorm  re_D re_M re_Mtime , n($NumPat)  cov(DTrue)  $reseed
mkmat re_M re_Mtime, matrix(RanEff)  // Each row of matrix RanEff corresponds to one simulated patient
*Random coefficients are computed
*A random coefficient is what is usually called a random effect plus its corresponding fixed effect.
quietly generate re_M_c=$b4True+re_M
quietly generate re_Time_c=$b7True+re_Mtime
*The errors are simulated
quietly drawnorm  `errors', n($NumPat)  cov(RTrue)  $erseed // errors are generated
mkmat `errors', matrix(Errors)     // Each row of matrix Errors contains errors for corresponding patient in
RanEff
*Columns of matrix MatbTrue will contain the true fixed effects repeatedly
matrix MatbTrue=bTrue
forvalues i=2/$NumPat {
matrix MatbTrue=MatbTrue,bTrue
}
*Matrix of responses is computed
*Each column of Y contains the simulated responses of the patient in corresponding column of RanEff'
matrix Z2=Z[1..rowsof(Z)-1, 2..3]
matrix X2=X[1..rowsof(X)-1, 4..8]
matrix Y=Z2*RanEff'+X2*MatbTrue+Errors'
```

```
matrix colnames Y=Patient
matrix rownames Y=PainScore
end
```

2.7.6. TrueParam_Module4.ado.

(Stata ado program used by trueparam.)

```
*Variance covariance matrix of error terms
matrix R=I(rowsof(Z))  *  $VarErr
matrix R[rowsof(Z),rowsof(Z)]=0
*Columns of matrix MatbGLS will contain the estimated fixed effects B repeatedly
matrix MatbGLS=B
forvalues i=2/$NumPat {
matrix MatbGLS=MatbGLS,B
}
*Residual
matrix zero=0
matrix zero_pt=zero
forvalues i=2/$NumPat {
matrix zero_pt=zero_pt,zero
}
matrix Ynew=Y\zero_pt
matrix Res=Ynew-X*MatbGLS
matrix Res2=Res[1..rowsof(Res)-1, 1...]
matrix Res3=Res2\zero_pt
*The EB predictors are computed
matrix BLUP=D*Z'*inv(R+Z*D*Z')*Res3
matrix rownames BLUP=re1 re2 re3
```

```
*EB predictors are saved in database
matrix BLUPT=BLUP'
svmat BLUPT, names(col)
end
```

TrueParam_Module5.ado.

(Stata ado program used by trueparam.)

```
*True benefit is computed
*(True variance of error is in global macro VarErrTrue)
quietly generate double TrueBenef=100*(normal(  ($y-(re_M_c + $b5True*$age + $b6True*$depress +
re_Time_c *($PredOrig+$h) + $b8True*$depress*($PredOrig+$h) )   ) /sqrt($VarErrTrue) ) ///
                                                                    - normal(   ($y-(re_M_c
+ $b5True*$age + $b6True*$depress) )   /sqrt($VarErrTrue)    ))
label var TrueBenef "True benefit (x100)"
*Predicted benefit is computed
*(Computed with parameters in Table 1 of article)
*(Variance of error from model in Table 1 is in global macro $VarErr)
*The EB predictors of the pain score model random effects are in re2 re3.
quietly generate double PredBenef=100*(normal(  ($y-($b4+re2 + $b5*$age + $b6*$depress + ($b7+re3)
*($PredOrig+$h) + $b8*$depress*($PredOrig+$h) )   ) /sqrt($VarErr) ) ///
                                                                    -   normal(          ($y-
($b4+re2 + $b5*$age + $b6*$depress) )   /sqrt($VarErr)    ))
label var PredBenef "Predicted benefit (x100)"
*Individual bias is computed
quietly generate double Bias=PredBenef-TrueBenef
*Individual relative bias is computed
quietly summarize Bias, detail
```

```
global MeanBias=r(mean)  //Mean bias
quietly summarize TrueBenef, detail
global MeanTrueBenef=r(mean)
global RelBias=($MeanBias/$MeanTrueBenef)*100  // Relative bias
quietly correlate TrueBenef PredBenef
global Correlation=r(rho)
end
```